DOCUMENT RESUME

ED 189 104 TM 800 241

TITLE	Truth in Testing Act of 1979: The Education Testing Act of 1979. Hearings before the Subcommittee on Elementary, Secondary and Vocational Education of the Committee on Education and Labor, House of Representatives, Ninety-Sixth Congress, First Session on H.R. 3564 and H.R. 4949 (Washington, D.C., July 31, August 1, September 10, September 24, and October 10-11, 1979).
INSTITUTION	Congress of the U.S., Washington, D.C. House Committee on Education and Labor.
PUB DATE	80
NOTE	1,198p.
EDRS PRICE	MF09 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS	*College Entrance Examinations: Confidentiality: Cost Effectiveness: Disclosure: Federal Regulation: Government School Relationship: Higher Education: Information Dissemination: *Legislation: *Occupational Tests: Scores: *Standardized Tests: Test Coaching: Test Items
IDENTIFIERS	*Educational Testing Act 1979: *Truth in Testing Act 1979

ABSTRACT

Testimony and exhibits presented at subcommittee hearings held in Washington, D.C., from July 31, 1979 through October 11, 1979, are compiled with reference to two proposed bills concerned with standardized admissions tests. The first, the Truth in Testing Act of 1979, H.R. 3564, requires all educational admissions testing conducted through interstate commerce, and all occupational admissions testing (which affects commerce) to be conducted with sufficient notice of test subject matter and test results, and for other purposes. The second, the Educational Testing Act of 1979, H.R.4949, requires certain information be provided to individuals who take standardized educational admissions tests, and for other purposes. This document contains the statements and exhibits presented by 127 persons. (Author/CTM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

TRUTH IN TESTING ACT OF 1979; THE EDUCATIONAL TESTING ACT OF 1979

HEARINGS

BEFORE THE

SUBCOMMITTEE ON ELEMENTARY, SECONDARY, AND VOCATIONAL EDUCATION

OF THE

COMMITTEE ON EDUCATION AND LABOR
HOUSE OF REPRESENTATIVES

NINETY-SIXTH CONGRESS

FIRST SESSION

ON

H.R. 3564

TO REQUIRE ALL EDUCATIONAL ADMISSIONS TESTING CONDUCTED THROUGH INTERSTATE COMMERCE, AND ALL OCCUPATIONAL ADMISSIONS TESTING (WHICH AFFECTS COMMERCE) TO BE CONDUCTED WITH SUFFICIENT NOTICE OF TEST SUBJECT MATTER AND TEST RESULTS, AND FOR OTHER PURPOSES

AND

H.R. 4949

TO REQUIRE CERTAIN INFORMATION BE PROVIDED TO INDIVIDUALS WHO TAKE STANDARDIZED EDUCATIONAL ADMISSIONS TESTS, AND FOR OTHER PURPOSES

HEARINGS HELD IN WASHINGTON, D.C., ON
JULY 31; AUGUST 1; SEPTEMBER 10, 24; OCTOBER 10
AND 11, 1979

Printed for the use of the Committee on Education and Labor

ED189104

TM800 241

U S DEPARTMENT OF HEALTH
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

TRUTH IN TESTING ACT OF 1979; THE EDUCATIONAL TESTING ACT OF 1979

APR 1 80
RECEIVED
ERIC/TM

HEARINGS

BEFORE THE

SUBCOMMITTEE ON ELEMENTARY, SECONDARY AND VOCATIONAL EDUCATION

OF THE

COMMITTEE ON EDUCATION AND LABOR
HOUSE OF REPRESENTATIVES

NINETY-SIXTH CONGRESS

FIRST SESSION

ON

H.R. 3564

TO REQUIRE ALL EDUCATIONAL ADMISSIONS TESTING CONDUCTED THROUGH INTERSTATE COMMERCE, AND ALL OCCUPATIONAL ADMISSIONS TESTING (WHICH AFFECTS COMMERCE) TO BE CONDUCTED WITH SUFFICIENT NOTICE OF TEST SUBJECT MATTER AND TEST RESULTS, AND FOR OTHER PURPOSES

AND

H.R. 4949

TO REQUIRE CERTAIN INFORMATION BE PROVIDED TO INDIVIDUALS WHO TAKE STANDARDIZED EDUCATIONAL ADMISSIONS TESTS, AND FOR OTHER PURPOSES

HEARINGS HELD IN WASHINGTON, D.C., ON
JULY 31; AUGUST 1; SEPTEMBER 10, 24; OCTOBER 10
AND 11, 1979

Printed for the use of the Committee on Education and Labor

U.S. GOVERNMENT PRINTING OFFICE
54-834 O WASHINGTON : 1980

COMMITTEE ON EDUCATION AND LABOR

CARL D. PERKINS, Kentucky, *Chairman*

FRANK THOMPSON, Jr., New Jersey
JOHN BRADEMAS, Indiana
AUGUSTUS F. HAWKINS, California
WILLIAM D. FORD, Michigan
PHILLIP BURTON, California
JOSEPH M. GAYDOS, Pennsylvania
WILLIAM (BILL) CLAY, Missouri
MARIO BIAGGI, New York
IKE ANDREWS, North Carolina
PAUL SIMON, Illinois
EDWARD P. BEARD, Rhode Island
GEORGE MILLER, California
MICHAEL O. MYERS, Pennsylvania
AUSTIN J. MURPHY, Pennsylvania
TED WEISS, New York
BALTASAR CORRADA, Puerto Rico
DALE E. KILDEE, Michigan
PETER A. PEYSER, New York
EDWARD J. STACK, Florida
PAT WILLIAMS, Montana
WILLIAM R. RATCHFORD, Connecticut
RAY KOGOVSEK, Colorado
DON BAILEY, Pennsylvania

JOHN M. ASHBROOK, Ohio
JOHN N. ERLENBORN, Illinois
JOHN H. BUCHANAN, Jr., Alabama
JAMES M. JEFFORDS, Vermont
WILLIAM F. GOODLING, Pennsylvania
MICKEY EDWARDS, Oklahoma
E. THOMAS COLEMAN, Missouri
KEN KRAMER, Colorado
ARLEN ERDAHL, Minnesota
THOMAS J. TAUKE, Iowa
DANIEL B. CRANE, Illinois
JON HINSON, Mississippi
THOMAS E. PETRI, Wisconsin

Subcommittee on Elementary, Secondary, and Vocational Education

CARL D. PERKINS, Kentucky, *Chairman*

WILLIAM D. FORD, Michigan
IKE ANDREWS, North Carolina
GEORGE MILLER, California
AUSTIN J. MURPHY, Pennsylvania
BALTASAR CORRADA, Puerto Rico
DALE E. KILDEE, Michigan
PAT WILLIAMS, Montana
AUGUSTUS F. HAWKINS, California
MICHAEL O. MYERS, Pennsylvania
RAY KOGOVSEK, Colorado

WILLIAM F. GOODLING, Pennsylvania
JOHN H. BUCHANAN, Jr., Alabama
ARLEN ERDAHL, Minnesota
DANIEL B. CRANE, Illinois
JON HINSON, Mississippi
JOHN M. ASHBROOK, Ohio,
(Ex Officio)

CONTENTS

Page
Text of H.R. 3564 ... 2
Text of H.R. 4949 ... 8

Hearings held in Washington, D.C. on:

July 31, 1979 ... 1
August 1, 1979 ... 319
September 10, 1979 ... 391
September 24, 1979 ... 531
October 10, 1979 ... 703
October 11, 1979 ... 821

Statement of—

Berendzen, Richard, president, American University, representing the American Council on Education, accompanied by Sheldon Steinbach and Douglas Whitney of the American Council on Education ... 536
Bond, Dr. Lloyd, research associate, learning research and development center, University of Pittsburgh ... 897
Cahow, Dr. Clark R., associate professor, history and assistant provost, Duke University, consortium on financing higher education ... 905
Cameron, Donald, assistant executive director for programs, National Education Association ... 718
Casteen, Dr. John T., III, dean of undergraduate admissions, University of Virginia ... 765
Clawson, Dr. D. Kay, dean, College of Medicine, University of Kentucky ... 863
Cooper, Dr. John A. D., president, Association of American Medical Colleges ... 70
DeWinter, Urbain J., director of admissions, College of Arts, Cornell University ... 532
Drexeler, Kenneth, president, student government, North Senior High School, Great Neck, N.Y ... 900
Erwin, Frank, member, advisory panel on declining test scores representing American Society for Personnel Administration ... 772
Fleishman, Dr. Edwin, past president, division of industrial and organizational psychology, American Psychological Association, president, Advanced Research Resources Organization ... 791
Gastfriend, David, chairman, Medical Education Committee, American Medical Students Association ... 868
Geer, William, Jr., principal, Francis W. Parker School, Chicago, Ill ... 911
Gibbons, Hon. Sam, a Representative in Congress from the State of Florida ... 31
Gould, Clifford, Deputy Director, Federal Personnel and Compensation Division, U.S. General Accounting Office ... 704
Goldstine, Sheldon R., Cleveland, Ohio ... 625
Graham, James W., assistant secretary, Council on Dental Education, American Dental Association ... 331
Haber, Jon, consultant on standardized testing, U.S. Student Association, accompanied by Andre Burnett, director, National Third World Student Coalition ... 358
Hall, Dr. Judy E., president-elect, American Association of State Psychology Boards ... 796
Halperin, Senator Donald, New York State Senate ... 329

Statement of—Continued

Page
Haney, Dr. Walter, testing expert, accompanied by Allan Nairn 67
Hargadon, Fred, chairman, board of trustees, the College Board 60
Herndon, Terry, executive director, National Education Association 32
Howe, Dr. Harold, II, vice president for education and public policy, the Ford Foundation 821
Jackson, Dr. George, National Association of Black Psychologists, accompanied by Alvis Adair 860
Keller, Dr. Gary, dean, graduate school, Eastern Michigan University, representing the Hispanic Higher Education Coalition 828
LaValle, Senator Kenneth, New York State Senate 323
Lennon, Roger T. assistant to the chairman in measurement, Harcourt Brace Jovanovich, Inc 365
Loewen, James W., director, research center for national policy review, School of Law, Catholic University of America 545
Newsom, Lionel, member, board of directors, National Association for Equal Opportunity in Higher Education, and president, Central State University (Ohio) 621
Perrone, Vito, dean, School of Education, University of North Dakota 449
Pottinger, Paul S., executive director, National Center for the Study of Professions 472
Rever, Philip, director, Washington office, American College Testing Program, accompanied by Richard Ferguson, vice president of ACT's research and development division 419
Rice, Lois D., vice president, College Board 60
Rivera, Dr. Alvin, Hispanic Higher Education Coalition 828
Ruud, Millard H., executive director, Association of American Law Schools 335
Shanker, Albert, president, American Federation of Teachers, AFL-CIO 855
Simmons, Althea T. L., director, Washington Bureau of the National Association for the Advancement of Colored People 555
Smith, I. Leon, director, professional examination service 642
Snyder, Frank, McGraw-Hill, representing Association of American Publishers, accompanied by Roy H. Millenson, AAP, Washington office 459
Solomon, Robert J., executive vice president, acting president, educational testing service 77
Sparling, Virginia, president, National Congress of Parents and Teachers 392
Weiss, Hon. Ted, a Representative in Congress from the State of New York . 39

Prepared statements, letters, supplemental materials, et cetera—

Berendzen, Dr. Richard, university provost and president-elect, the American University, on behalf of the American Association of State Colleges and Universities; American Council on Education; Association of American Universities; Association of Catholic Colleges and Universities; Council of Graduate Schools of the United States; National Association for Equal Opportunity in Higher Education; National Association of Independent Colleges and Universities; and National Association of State Universities and Land-Grant Colleges, prepared statement of 541
Cahow, Dr. Clark R., associate professor, history and assistant provost, Duke University, prepared statement of 908
Cameron, Don, assistant executive director, National Education Association, prepared statement with appendix 720
Casteen, Dr. John T., III, dean of undergraduate admissions, University of Virginia:
  Letter enclosing information referred to 819
  Prepared statement of 768
Chisholm, Hon. Shirley, a Representative in Congress from the State of New York, prepared statement of 24
Clawson, Dr. D. Kay, dean, College of Medicine, University of Kentucky, prepared statement of 866
Cooper, Dr. John A. D., president, Association of American Medical Colleges:
  Appendix I to statement 73
  "Educational Testing Service, Finances—Where the Money Comes From and Where It Goes," paper entitled 237
  "ETS 1978, Annual Report," pamphlet entitled 210

Prepared statements, letters, supplemental materials, etc.—Continued

Cooper, Dr. John A. D.—Continued — Page

- Letter from Chairman Perkins, dated November 9, 1979, requesting additional information 193
- Letter to Chairman Perkins, dated August 10, 1979 191
- Letter to Chairman Perkins, dated November 14, 1979 194
- Letter to Chairman Perkins, dated November 28, 1979, enclosing information requested 204
- Letter to House Subcommittee on Elementary, Secondary, and Vocational Education, dated November 13, 1979 195
- "Program Research Progress Report," August 1979 249

Drexeler, Kenneth, president, student government, North Senior High School, Great Neck, N.Y.:

- Editorial, "The Medical Colleges Flunk This Test," from Newsday, July 23, 1979 904
- "Truth in Testing," article from the New York Times, July 24, 1979 ... 904

Erwin, Frank W., The American Society for Personnel Administration, statement on behalf of 779

Gastfriend, David, chairman, Medical Education Committee, American Medical Students Association:

- Exhibit A, "Academic and Personal Predictors of Clinical Success in Medical School" 873
- Exhibit B, "Predicting Supervisor, Peer, and Self Ratings of Intern Performance" 882
- Exhibit C, "Economic and Racial Disadvantage as Reflected in Traditional Medical School Selection Factors" 886

Geer, William, Jr., principal, Francis W. Parker School, Chicago, Ill., prepared statement with attachments 914

Coldstine, Sheldon R., Cleveland, Ohio, letter to Chairman Perkins, dated September 24, 1979 632

Gould, Clifford I., Deputy Director, Federal Personnel and Compensation Division, U.S. General Accounting Office:

- Prepared statement of 706
- "Recommendations to the Director, Office of Personnel Management," article entitled 717

Haber, Jon, representing the U.S. Student Association, accompanied by Andre Burnett, director, National Third World Student Coalition, U.S. Student Association, prepared statement of 352

Hall, Dr. Judy E., president-elect, American Association of State Psychology Boards, "Examination for Professional Practice in Psychology," article entitled 801

Herndon, Terry, executive director, National Education Association, prepared statement of 321

Howe, Dr. Harold, II, vice president for education and public policy, the Ford Foundation:

- Prepared statement of 825
- Prepared statement of Dr. Kenneth B. Clark, distinguished professor emeritus, City College of the City University of New York 822

Keller, Dr. Gary, dean, graduate school, Eastern Michigan University, representing the Hispanic Higher Education Coalition, prepared statement of 833

LaValle, Senator Kenneth, New York State Senate, prepared statement of . 327

Loewen, Dr. James W., director of research, Center for National Policy Review, School of Law, Catholic University of America:

- Additional statement of, dated September 24, 1979 552
- Prepared statement of, dated August 15, 1979 548

Magazine and newspaper articles—

- Boston Globe:
  - "Open Test Law Applauded Here," dated July 16, 1979 667
  - "Testing the Test-Makers," dated July 19, 1979 671
- Chicago Sun-Times: "Demystify College Entrance Exams," dated July 18, 1979 666
- Higher Education Daily: "Howard University Study Cites Limits of Standardized Tests," dated August 29, 1979 675

Prepared statements, letters, supplemental materials, etc.—Continued

Magazine and newspaper articles—Continued Page

New York Times:

"Test-Coaching: Is It Worth It?," dated September 9, 1979 680
"Test Publishers See Dropping of Exams," dated October 8, 1979.. 754
"Truth in Testing," dated July 24, 1979 672
"'Truth-in-Testing' Measure Focusing on a Key Industry," dated July 5, 1979 669
New York Times magazine: "The American Way of Testing," dated September 2, 1979 676

The New Republic:

"ETS Has Good Reason to Fear New York's New 'Truth-in-Testing' Law, Evening the Score," dated August 25, 1979 672
Why Your Daughter Didn't Get Into Law School, ETS's Star Chamber," dated February 5, 1977 668
Washington Post: "Scoring the Admissions Tests," dated August 27, 1974 507
Miller, Hon. George, a Representative in Congress from the State of California, prepared statement of 28
National Consortium on Testing, list of members, dated June 1, 1979 494
Perrone, Vito, center for teaching and learning, University of North Dakota, appendix to the prepared statement of, dated October 30, 1978 455
Pike, Dr. Lewis, testimony before the New York State Senate and Assembly Higher Education Committees, May 9, 1979, Albany, N.Y 683
Pottinger, Paul S., Ph. D., director National Center for the Study of Professions, Washington, D.C., prepared statement of 477
Rever, Philip, director, Washington office, American College Testing Program, statement of 423
Ruud, Millard H., executive director, Association of American Law Schools, prepared statement of 337
Shanker, Albert, president, American Federation of Teachers, AFL-CIO, prepared statement of 852

Simmons, Althea T. L., director, Washington Bureau of the National Association for the Advancement of Colored People, accompanied by Dr. Anderson J. Franklin, NAACP Advisory Committee on Testing and professor of psychology and director, center for academic skills, City College of CUNY:

"NAACP Report on Minority Testing" 582
Prepared statement of 560
Smith, Dr. I. Leon, professional examination service, testimony of 647
Snyder, Frank, publisher, CTB/McGraw-Hill, Monterey, Calif., testimony on behalf of Association of American Publishers, with attachment 466

Solomon, Robert J., executive vice president, acting president, educational testing service:

Attachment 1, agreement dated July 1, 1974 between the law school admission council and educational testing service 88
Attachment 2, members of ETS board of trustees 1974-79 94
Attachment 3, (retained in subcommittee files)
Attachment 4, historical change of fees versus inflation 95
Attachment 5, comments on H.R. 4949 and H.R. 3564 100
Attachment 6, some implications on test development costs and test availability of national test disclosure 107
Attachment 7, principles, policies and procedural guidelines regarding ETS products and services 110
Sparling, Virginia, president, National Congress of Parents and Teachers, "The Testing Maze, An Evaluation of Standardized Testing in America," paper prepared and published by the national PTA 395
Wallach, Michael A., article entitled "Tests Tell Us Little About Talent," from American Scientist, January-February 1976 950
Weiss, Hon. Ted, a Representative in Congress from the State of New York, prepared statement of 43

APPENDIX

Allen, William E., D.D.S., chairman, council on legislation, American Dental Association, Washington, D.C., letter to Chairman Perkins, dated October 11, 1979 1141

Page

Ambach, Gordon M., president, the University of the State of New York and commissioner of education:
  Letter to Mr. Chairman, with enclosures, dated October 17, 1979 ... 966
  Prepared statement of ... 959
American Educational Research Association, prepared statement of the, with attachment ... 1045
American Nurses' Association, prepared statement of the ... 1079
American Society of Clinical Pathologists, prepared statement of the, with attachments ... 1014
Bethune, Hon. Ed, a Representative in Congress from the State of Arkansas, letter to Chairman Perkins, with enclosures, dated October 31, 1979 ... 1176
Cave, Carolyn, Ph. D., RRA, executive director, American Medical Record Association, Chicago, Ill., letter to Chairman Perkins, dated October 30, 1979 .. 1174
Corbett, William D., Waltham, Mass., testimony of ... 1182
Down, A. Graham, executive director, Council for Basic Education, Washington, D.C., letter to Chairman Perkins, dated August 29, 1979 ... 1120
Eckenhoff, James E., M.D., dean, Northwestern University, Chicago, Ill., letter to members of the Education and Labor Committee, dated August 16, 1979 .. 1113
Federation of Associations of Schools of the Health Professions, statement of the ... 999
Floyd, Edwin E., dean, University of Virginia, Charlottesville, letter to Committee on Education and Labor, dated August 17, 1979 ... 1117
Ford, Hon. William D., a Representative in Congress from the State of Michigan, letter to Chairman Perkins, with enclosure, dated October 24, 1979 ... 1168
Galvin, C. R., director, Real Estate Commission, Des Moines, Iowa, letter to Chairman Perkins, dated September 10, 1979 ... 1123
Glaser, Robert, president, Division of Evaluation and Measurement, American Psychological Association, prepared statement of ... 1080
Green, Bert F., professor of psychology, the Johns Hopkins University, Baltimore, Md., letter to Congress of the United States, dated October 17, 1979 ... 1143
Gustafson, Delwin K., president, National Association of College Admissions Counselors, Skokie, Ill., letter to Chairman Perkins, dated October 23, 1979 . 1165
Harmon, George M., president, Millsaps College, Jackson, Miss., letter to Congressman Hinson, dated September 7, 1979 ... 1122
Herndon, Robert M., M.D., director, center for brain research, University of Rochester Medical Center, Rochester, N.Y., letter to Chairman Perkins, dated August 17, 1979 ... 1116
Hill, Marilyn J., associate director, placement coordinator, student futures and placement, Carleton College, Northfield, Minn., letter to Congressman Erdahl, dated September 14, 1979 ... 1129
Jackson, William F., D.V.M., president, American Veterinary Medical Association, prepared statement of ... 1084
Jervey, Dr. Harold E., Jr., executive director-secretary, Federation of State Medical Boards of the United States, Inc., prepared statement of ... 1065
Kernis, Marten M., Ph. D., deputy executive dean, University of Illinois College of Medicine, Chicago, Ill., letter to Chairman Perkins, dated August 16, 1979... 1115
Kerr, Clark, chairman, Carnegie Council on Policy Studies in Higher Education, Berkeley, Calif., letter to Chairman Perkins, dated August 14, 1979...... 1108
Kiernan, Dr. Owen B., executive director, National Association of Secondary School Principals, prepared statement of ... 1040
Leymaster, Glen R., M.D., executive director, American Board of Medical Specialties, Evanston, Ill., letter to Chairman Perkins, with attachment, dated October 9, 1979 ... 1133
Madigan, Hon. Edward R., a Representative in Congress from the State of Illinois, letter to Chairman Perkins, with enclosure, dated September 19, 1979 1131
Mahaffey, Fred T., Pharm. D., executive director, National Association of Boards of Pharmacy, Chicago, Ill., letter to Chairman Perkins, dated July 20, 1979 ... 1106
Mann, Wallace V., Jr., D.M.D., dean, School of Dentistry, University of Mississippi Medical Center, Jackson, Miss., letter to Congressman Hinson, dated September 12, 1979 ... 1125
Maraldo, Pamela, public affairs director, National League for Nursing, New York, N.Y., letter to Chairman Perkins enclosing a statement, dated December 18, 1979 ... 1183
Marquis, Harold L., acting dean, Emory University School of Law, Atlanta, Ga., letter to Chairman Perkins, dated November 5, 1979 ... 1180

Page

Mast, Kathleen A., RDH, president, American Dental Hygienists' Association, Washington, D.C., letter to Chairman Perkins, dated August 20, 1979 1118

McCombs, Robert M., Ph. D., interim dean, Eastern Virginia Medical School, Norfolk, Va., letter to Chairman Perkins, dated August 15, 1979 1112

National Board of Medical Examiners, Philadelphia, Pa., prepared statement of 1005

National Council of Architectural Registration Boards, prepared statement of the, with attachment 1095

National League for Nursing, New York, N.Y., prepared statement of 1075

Nobles, Lewis, president, Mississippi College, Clinton, Miss., letter to Congressman Hinson, dated September 13, 1979 1127

Pelczar, Michael J., Jr., president, the Council of Graduate Schools in the United States, Washington, D.C., letter to Chairman Perkins, with enclosures, dated October 19, 1979 1145

Scerba, John R., Ph. D., North Miami, Fla., letter to Chairman Perkins, dated October 24, 1979 1172

Shannon, David A., vice president and provost, University of Virginia, Charlottesville, letter to Chairman Perkins, dated August 14, 1979 1110

Sjogren, Dr. Cliff, director of admissions, University of Michigan—Ann Arbor, prepared statement of 1001

Stockton, Jack J., D.V.M., president, Association of American Veterinary Medical Colleges, prepared statement of 1093

Stratman, Dr. David G., director, Campaign for the Future, prepared statement of 1062

White, Donald J., dean, Boston College, Chestnut Hill, Mass., letter to Chairman Perkins, dated October 22, 1979 1164

Willcox, Alfred B., executive director, the Mathematical Association of America, prepared statement of 1086

Wright, Grace H., Ph. D., research psychologist, director, validation program, New York State Department of Civil Service, letter to Chairman Perkins enclosing several attachments, dated December 20, 1979 1185

Zytowski, Donald G., Ed. D., counseling psychologist and professor of psychology, Iowa State University, on behalf of the American Personnel and Guidance Association, prepared statement of 1054

TRUTH IN TESTING ACT OF 1979; THE EDUCATIONAL TESTING ACT OF 1979

TUESDAY, JULY 31, 1979

House of Representatives,
Subcommittee on Elementary, Secondary,
and Vocational Education,
Committee on Education and Labor,
Washington, D.C.

The subcommittee met, pursuant to notice, at 9:30 a.m., in room 2175, Rayburn House Office Building, Hon. Carl D. Perkins (chairman of the committee) presiding.

Members present: Representatives Perkins, Hawkins, Miller, Weiss, Kildee, Williams, Goodling, and Erdahl.

Staff present: John F. Jennings, counsel; Martin L. LaVor, minority senior legislative associate.

[The complete text of H.R. 3564 and H.R. 4949 follows:]

96TH CONGRESS
1ST SESSION

H. R. 3564

To require all educational admissions testing conducted through interstate commerce, and all occupational admissions testing (which affects commerce) to be conducted with sufficient notice of test subject matter and test results, and for other purposes.

IN THE HOUSE OF REPRESENTATIVES

APRIL 10, 1979

Mr. GIBBONS introduced the following bill; which was referred to the Committee on Education and Labor

---

A BILL

To require all educational admissions testing conducted through interstate commerce, and all occupational admissions testing (which affects commerce) to be conducted with sufficient notice of test subject matter and test results, and for other purposes.

Be it enacted by the Senate and House of Representatives of the United States of America in Congress assembled, That this Act may be cited as the "Truth in Testing Act of 1979".

SEC. 2. As used in this Act—

(1) the term "educational admissions test" means any test of aptitude or knowledge which—

(A) is administered to individuals in two or more States,

(B) affects or is conducted or distributed through any medium of interstate commerce, and

(C) is used as part or all of the basis for admitting or denying admission to an individual to any institution of higher education;

(2) the term "occupational admissions test" means any test which is used as part or all of the basis for admitting or denying admission to an individual to any occupation in or affecting interstate commerce;

(3) the term "test" includes any aptitude or achievement examination, whether written or oral, and includes any objective multiple choice, machine scored, essay, practical, performance, or demonstration examination;

(4) the term "test score" means the numerical value given to the test subject's performance on any test;

(5) the term "person" includes individuals, corporations, companies, associations, firms, partnerships, societies, joint stock companies, and agencies and instrumentalities of States and local governments; and

(6) the term "institution of higher education" has the meaning set forth in section 1201(a) of the Higher Education Act of 1965 (20 U.S.C. 1141(a)).

SEC. 3. The Congress hereby finds and declares that—

(1) testing of scholastic aptitudes and achievements has become a principal factor in the admission of individuals to public, as well as to private, institutions of higher education and that therefore equal opportunity under the law requires that that testing be conducted in a manner which will ensure the equal rights and fair treatment of such individuals;

(2) testing of skills for entry into an occupation, whether of a professional, craft, or trade nature, is a critical factor governing the free flow of individual skills in interstate commerce and seriously affects the Nation's capability for economic growth; and

(3) the rights of individuals and the national interests can be protected without adversely affecting the proprietary interest of any entity administering tests by simple requirements governing proper prior notice to individuals of the subject matter to be tested and proper subsequent notice of test results and their uses.

SEC. 4. It is the purpose of this Act to prohibit the conducting of educational and occupational admissions tests unless such tests are administered in a manner to protect the rights of the individuals tested and to grant a Federal cause of action to any individual adversely affected by the administration of any such test in violation of this Act.

SEC. 5. It is unlawful for any person to administer any educational or occupational admissions test to any individual unless such test is administered in accordance with the requirements of section 6 of this Act.

SEC. 6. (a) Each applicant to take any educational or occupational admissions test shall be provided with a written notice which shall contain—

(1) a detailed description of the area of knowledge or the type of aptitude that the test attempts to analyze;

(2) in the case of a test of knowledge, a detailed description of the subjects to be tested;

(3) the margin of error or the extent of reliability of the test, determined on the basis of experimental uses of the test and, where available, actual usage;

(4) the manner in which the test results will be distributed by the testing entity to the applicant and to other persons; and

(5) a statement of the applicant's rights under subsection (b) of this section to obtain test results and related facts.

(b) Each individual who takes any educational or occupational admissions test shall, at the request of the test subject, promptly upon completion of scoring of such test, be notified of—

(1) the individual's specific performance in each of the subject or aptitude areas tested;

(2) how that specific performance ranked in relation to the other individuals and how the individual ranked on total test performance;

(3) the score required to pass the test for admission to such occupation or the score which is generally required for admission to institutions of higher education;

(4) any further information which may be obtained by the individual on request.

(c) No educational or occupational admissions test which tests knowledge or achievement (rather than aptitude) shall be graded (for purposes of determining the score required to pass the test for admission) on the basis of the relative distribution of scores of other test subjects.

SEC. 7. (a) Whenever any person has administered or there are reasonable grounds to believe that any person is about to administer any educational or occupational admissions test in violation of this Act, a civil action for preventive relief, including an application for a permanent or temporary injunction, restraining order, or other order, may be instituted by the individual or individuals aggrieved. Upon application by the complainant and in such circumstances as the court may deem just, the court may appoint an attorney for such complainant and may authorize the commencement of civil action without payment of fees, costs, or security.

(b) In any action commenced pursuant to this section, the court, in its discretion, may allow the prevailing party, other than the United States, a reasonable attorney's fee as part of the costs.

(c) The district courts of the United States shall have jurisdiction of proceedings instituted pursuant to this Act and shall exercise the same without regard to whether the aggrieved party shall have exhausted any administrative or other remedies that may be provided by law.

SEC. 8. This Act shall be effective with respect to any test administered on or after January 1, 1979.

96TH CONGRESS
1ST SESSION

H. R. 4949

To require certain information be provided to individuals who take standardized educational admissions tests, and for other purposes.

IN THE HOUSE OF REPRESENTATIVES

JULY 24, 1979

Mr. WEISS (for himself, Mrs. CHISHOLM, and Mr. MILLER of California) introduced the following bill; which was referred to the Committee on Education and Labor

---

A BILL

To require certain information be provided to individuals who take standardized educational admissions tests, and for other purposes.

Be it enacted by the Senate and House of Representatives of the United States of America in Congress assembled,

SHORT TITLE

SECTION 1. This Act may be cited as the "Educational Testing Act of 1979".

FINDINGS AND PURPOSE

SEC. 2. (a) The Congress of the United States finds that—

(1) education is fundamental to the development of individual citizens and the progress of the Nation as a whole;

(2) there is a continuous need to ensure equal access for all Americans to educational opportunities of a high quality;

(3) standardized tests are a major factor in the admission and placement of students in postsecondary education and also play an important role in individuals' professional lives;

(4) there is increasing concern among citizens, educators, and public officials regarding the appropriate uses of standardized tests in the admissions decision of postsecondary education institutions;

(5) the rights of individuals and the public interest can be assured without endangering the proprietary rights of the testing agencies; and

(6) standardized tests are developed and administered without regard to State boundaries and are utilized on a national basis.

(b) It is the purpose of this Act—

(1) to ensure that test subjects and persons who use test results are fully aware of the characteristics, uses, and limitations of standardized tests in postsecondary education admissions;

(2) to make available to the public appropriate information regarding the procedures, development, and administration of standardized tests;

(3) to protect the public interest by promoting more knowledge about appropriate use of standardized test results and by promoting greater accuracy, validity, and reliability in the development, administration, and interpretation of standardized tests; and

(4) to encourage use of multiple criteria in the grant or denial of any significant educational benefit.

INFORMATION TO TEST SUBJECTS AND POSTSECONDARY EDUCATIONAL INSTITUTIONS

SEC. 3. (a) Each test agency shall provide to any test subject in clear and easily understandable language, along with the registration form for a test, the following information:

(1) The purposes for which the test is constructed and is intended to be used.

(2) The subject matters included on such test and the knowledge and skills which the test purports to measure.

(3) Statements designed to provide information for interpreting the test results, including explanations of the test, and the correlation between test scores and future success in schools and, in the case of tests used

for postbaccalaureate admissions, the standard error of measurement and the correlation between test scores and success in the career for which admission is sought.

(4) Statements concerning the effects on and uses of test scores, including—

(A) if the test score is used by itself or with other information to predict future grade point average, the extent, expressed as a percentage, to which the use of this test score improves the accuracy of predicting future grade point average, over and above all other information used; and

(B) a comparison of the average score and percentiles of test subjects by major income groups; and

(C) the extent to which test preparation courses improve test subjects' scores on average, expressed as a percentage.

(5) A description of the form in which test scores will be reported, whether the raw test scores will be altered in any way before being reported to the test subject, and the manner, if any, the test agency will use the test score (in raw or transformed form) by itself or together with any other information about the test subject to predict in any way the subject's future aca-

demic performance for any postsecondary educational institution.

(6) A complete description of any promises or covenants that the test agency makes to the test subject with regard to accuracy of scoring, timely forwarding or score reporting, and privacy of information (including test scores and other information), relating to the test subjects.

(7) The property interests of the test subject in the test results, if any, the duration for which such results will be retained by the test agency, and policies regarding storage, disposal, and future use of test scores.

(8) The time period within which the test subject's test score will be completed and mailed to the test subject and the time period within which such scores will be mailed to test score recipients designated by the test subject.

(9) A description of special services to accommodate handicapped test subjects.

(10) Notice of (A) the information which is available to the test subject under section 5(a)(2), (B) the rights of the test subject under section 6, and (C) the procedure for appeal or review of a test score by the test agency.

(b) Any institution which is a test score recipient shall be provided with the information required by subsection (a). The test agency shall provide such information with respect to any test prior to or coincident with the first reporting of a test score or scores for that test to a recipient institution.

(c) The test agency shall immediately notify the test subject and the institutions designated as test score recipients by the test subject if the test subject's score is delayed ten calendar days beyond the time period stated under subsection (a)(8) of this section.

REPORTS AND STATISTICAL DATA AND OTHER INFORMATION

SEC. 4. (a)(1) In order to further the purposes of this Act, the following information shall be provided to the Commissioner by the test agency:

(A) Any study, evaluation, or statistical report pertaining to a test, which a test agency prepares or causes to be prepared, or for which it provides data. Nothing in this paragraph shall require submission of any reports or documents containing information identifiable with any individual test subject. Such information shall be deleted or obliterated prior to submission to the Commissioner.

(B) If one test agency develops or produces a test and another test agency sponsors or administers the same test, a copy of their contract for services shall be submitted to the Commissioner.

(2) All data, reports, or other documents submitted pursuant to this section will be considered to be records for purposes of section 552(a)(3) of title 5, United States Code.

(b) Within one year of the effective date of this Act, the Commissioners shall report to Congress concerning the relationship between the test scores of test subjects and income, race, sex, ethnic, and handicapped status. Such report shall include an evaluation of available data concerning the relationship between test scores and the completion of test preparation courses.

PROMOTING A BETTER UNDERSTANDING OF TESTS

SEC. 5. (a) In order to promote a better understanding of standardized tests and stimulate independent research on such tests, each test agency—

(1) shall, within thirty days after the results of any standardized test are released, file or cause to be filed in the office of the Commissioner—

(A) a copy of all test questions used in calculating the test subject's raw score;

(B) the corresponding acceptable answers to those questions; and

(C) all rules for transferring raw scores into those scores reported to the test subject and postsecondary educational institutions together with an explanation of such rules; and

(2) shall, after the test has been filed with the Commissioner and upon request of the test subject, send the test subject—

(A) a copy of the test questions used in determining the subject's raw score;

(B) the test subject's individual answer sheet together with a copy of the correct answer sheet to the same test with questions counting toward the test subject's raw score so marked; and

(C) a statement of the raw score used to calculate the scores already sent to the test subject if such request has been made within ninety days of the release of the test score to the test subject.

The test agency may charge a nominal fee for sending out such information requested under paragraph (2) not to exceed the marginal cost of providing the information.

(b) This section shall not apply to any standardized test for which it can be anticipated, on the basis of past experience (as reported under section 7(2) of this Act), will be administered to fewer than five thousand test subjects nationally over a testing year.

(c) Documents submitted to the Commissioner pursuant to this section shall be considered to be records for purposes of section 552(a)(3) of title 5, United States Code.

PRIVACY OF TEST SCORES

SEC. 6. The score of any test subject, or any altered or transferred version of the score identifiable with any test subject, shall not be released or disclosed by the test agency to any person, organization, association, corporation, postsecondary educational institution, or governmental agency or subdivision unless specifically authorized by the test subject as a score recipient. A test agency may, however, release all previous scores received by a test subject to any currently designated test score recipient. This section shall not be construed to prohibit release of scores and other information in a form which does not identify the test subject for purposes of research leading to studies and reports primarily concerning the tests themselves.

TESTING COSTS AND FEES TO STUDENTS

SEC. 7. In order to ensure that tests are being offered at a reasonable cost to test subjects, each test agency shall report the following information to the Commissioner:

(1) Before March 31, 1981, or within ninety days after it first becomes a test agency, whichever is later, the test agency shall report the closing date of its testing year. Each test agency shall report any change in

the closing date of its testing year within ninety days after the change is made.

(2) For each test program, within one hundred and twenty days after the close of the testing year the test agency shall report—

(A) the total number of times the test was taken during the testing year;

(B) the number of test subjects who have taken the test once, who have taken it twice, and who have taken it more than twice during the testing year;

(C) the number of refunds given to individuals who have registered for, but did not take, the test;

(D) the number of test subjects for whom the test fee was waived or reduced;

(E) the total amount of fees received from the test subjects by the test agency for each test program for that test year;

(F) the total amount of revenue received from each test program; and

(G) the expenses to the test agency of the tests, including—

(i) expenses incurred by the test agency for each test program;

(ii) expenses incurred for test development by the test agency for each test program; and

(iii) all expenses which are fixed or can be regarded as overhead expenses and not associated with any test program or with test development;

(3) If a separate fee is charged test subjects for admissions data assembly services or score reporting services, within one hundred and twenty days after the close of the testing year, the test agency shall report—

(A) the number of individuals registering for each admissions data assembly service during the testing year;

(B) the number of individuals registering for each score reporting service during the testing year;

(C) the total amount of revenue received from the individuals by the test agency for each admissions data assembly service or score reporting service during the testing year; and

(D) the expenses to the test agency for each admissions data assembly service or score reporting service during the testing year.

REGULATIONS AND ENFORCEMENT

SEC. 8. (a) The Commissioner shall promulgate regulations to implement the provisions of this Act within one hundred and twenty days after the effective date of this Act. The failure of the Commissioner to promulgate regulations shall not prevent the provisions of this Act from taking effect.

(b) Any test agency that violates any clause of any provision of this Act shall be liable for a civil penalty not to exceed $2,000 for each violation.

(c) If any provision of this Act shall be declared unconstitutional, invalid, or inapplicable, the other provisions shall remain in effect.

DEFINITIONS

SEC. 9. For purposes of this Act—

(1) the term "admissions data assembly service" means any summary or report of grades, grade point averages, standardized test scores, or any combination of grades and test scores, of an applicant used by any postsecondary educational institution in its admissions process;

(2) the term "Commissioner" means the Commissioner of Education;

(3) the term "postsecondary educational institution" means any institution providing a course of study

beyond the secondary school level and which uses standardized tests as a factor in its admissions process;

(4) the term "score reporting service" means the reporting of a test subject's standardized test score to a test score recipient by a testing agency;

(5) the term "standardized test" or "test" means—

(A) any test that is used, or is required, for the process of selection for admission to postsecondary educational institutions or their programs, or

(B) any test used for preliminary preparation for any test that is used, or is required, for the process of selection for admission to postsecondary educational institutions or their programs,

which affects or is conducted or distributed through any medium of interstate commerce, but such term does not include any test designed solely for nonadmission placement or credit-by-examination or any test developed and administered by an individual school or institution for its own purposes only;

(6) the term "test agency" means any person, organization, association, corporation, partnership, or in-

dividual which develops, sponsors, or administers a standardized test;

(7) the term "test preparation course" means any curriculum, course of study, plan of instruction, or method of preparation given for a fee which is specifically designed or constructed to prepare a test subject for, or to improve a test subject's score on, a standardized test;

(8) the term "test program" means all the administrations of a test of the same name during a testing year;

(9) the term "test score" means the value given to the test subject's performance by the test agency on any test, whether reported in numerical, percentile, or any other form.

(10) the term "test score recipient" means any person, organization, association, corporation, postsecondary educational institution, or governmental agency or subdivision to which the test subject requests or designates that a test agency reports his or her score;

(11) the term "test subject" means an individual to whom a test is administered; and

(12) the term "testing year" means the twelve calendar months which the test agency considers either its operational cycle or its fiscal year.

EFFECTIVE DATE

SEC. 10. This Act shall take effect one hundred and eighty days after the date of its enactment.

Chairman **Perkins**. The Subcommittee on Elementary, Secondary, and Vocational Education is conducting hearings today and tomorrow on two bills related to standardized testing.

The first bill, H.R. 3564, the Truth in Testing Act of 1979, was introduced by Congressman Sam Gibbons. This legislation would affect educational and occupational admissions testing conducted through interstate commerce. The bill requires that certain information about test subject matter and test results be made available to persons taking these tests.

The second bill, H.R. 4949, introduced by Congressman Weiss, is known as the Educational Testing Act of 1979. This bill would cover all tests used for admission to postsecondary educational institutions and programs. The Weiss bill also requires that certain information about the nature of the test be supplied to test takers prior to the examination. Moreover, the legislation mandates disclosure of test questions and correct answers after an exam has been given upon request of the test-taker.

Both of these bills raise some very important issues regarding the rights of persons who take standardized tests. We have invited some very distinguished panels of Members of Congress and persons familiar with and involved in standardized testing to respond to these bills. I am sure the testimony today and tomorrow will be most helpful to the members as we consider this legislation.

I am delighted to welcome one of our most prominent Members of the House of Representatives, a former member of this committee, and a great legislator, the Honorable Sam Gibbons.

[Statements submitted by Representative Shirley Chisholm and Representative George Miller follow:]

PREPARED STATEMENT OF HON. SHIRLEY CHISHOLM, A REPRESENTATIVE IN CONGRESS FROM THE STATE OF NEW YORK

As one who has been involved in the education of children prior to and throughout my legislative career, I am particularly pleased to present my views on the H.R.4949, the Educational Testing Act of 1979, and to be associated with this legislation as one of its cosponsors. The impact that standardized tests have on the educational futures and careers of millions of Americans is not a subject of serious debate; standardized tests dictate to a very substantial degree the life choices of vast numbers of our citizens. In many cases, standardized tests may be used as a gatekeeping device for access to educational opportunities. For these reasons, the sponsors of the Educational Testing Act believe that it is imperative that we implement some uniform standard of openness which will serve to enhance our understanding of standardized testing and imbue a fundamental element of fairness in the testing process.

There have been a variety of objectives attributed to the sponsors and proponents of H.R.4949, along with much misinformation regarding the impact of the bill on the standardized testing process. As written, the bill will have no direct impact on the content of standardized tests or the uses of such tests in the admissions processes of educational institutions. The legislation can neither be construed as a tool to regulate the testing industry nor as an inappropriate

Federal foray into the admissions policies of colleges and universities. The legislation is a disclosure bill, no more, no less, and I believe that it is vital to opening up this important educational issue to broader research and public debate.

I want to emphasize that I have no delusions about what this legislation can accomplish. We are all aware of the problems and controversies that have surfaced surrounding the impact of standardized tests on minority students, and the great potential for cultural, racial and geographical bias written into test questions. However, I want to make it clear that in no way do I envision this legislation to be a placebo for the problems that confront minorities and other groups that may face inherent biases in the examination process.

For too long testing procedures and uses have remained shrouded in secrecy and ambiguity, an atmosphere that is not conducive to the educational interests of our children. We have heard substantial testimony from experts in the testing field that test questions and answers, validity studies, cultural bias analyses, and other data vital in ascertaining the appropriateness and impact of standardized test on non-white and non-urban students have often been kept from the scrutiny of independent researchers, who seek to look beyond the body of knowledge currently available. Establishing a nationwide policy of openness through enactment of truth-in-testing legislation is a necessary first step to approaching the vital issues of concern to those who may find themselves adversely affected by a process which is presently void of accountability.

H.R. 4949 would enhance the elimination of biases in question selection not by government decree, but by enabling broader and more objective research into the process of test preparation and the criteria used in question selection. This kind of independent evaluation is almost impossible at present, because of the secrecy in which the industry shrouds its questions and answers, studies, and procedures. Our bill will guarantee access to essential material by requiring the companies to open their analyses of test results to public inspection.

The Subcommittee has heard ample testimony on the individual provisions of this legislation, and I will therefore be brief in my statement today. I would like to have the following article, published in the Washington Post, October 27, 1979, and authored by the three sponsors of H.R. 4949, entered into the record with my statement. This piece goes into a more detailed explanation of our objectives behind this legislative initiative and the importance of specific provisions.

The hearings that have been held to date on this legislation indicate that there is a dire need for members of this Congress to become more educated about the extent that testing affects the lives of their constituents, and the problems of extensive usage of standardized tests. I believe that the public debate that is beginning to develop around this bill as a result of Congressional action and subsequent media coverage is a vital and healthy debate. I look foward to the growth of this debate in both scope and interest as the legislative progress of this important issue continues.

* * * * *

Ted Weiss, Shirley Chisholm and George Miller

Tests Without Mystery

> *"Too often an individual's academic future is made or broken by a test result that may not even be accurate."*

Our proposed "truth-in-testing" legislation is not laying the foundation for a government-run testing system as The Washington Post (Aug. 27 editorial) and columnist William Raspberry (Sept. 10 and Oct. 1 commentary) contend.

They charge that our bill, the Educational Testing Act of 1979, would not correct admitted injustices and biases in the standardized exam process, but would instead inflict additional harm on the very people already victimized by the tests—primarily poor and minority students. Critics also maintain that the legislation is addressed to the needs of a special group and will only create a bigger market for the coaching schools that prepare students for the exams.

A careful look at our legislation's purposes and provisions should dispel these fears and correct misunderstandings. The Educational Testing Act is a disclosure measure that encourages the testing companies themselves to make needed reforms. It is not a regulatory intrusion or an attempted takeover by Big Brother. It will promote fairness and openness to the benefit of all test-takers.

In no way does the bill demand alterations in the examinations or in admission standards for academic institutions. The legislation in fact strengthens the reliability of standardized testing by helping to demystify the entire process and to establish a method of cross-checking scores.

The handful of tax-exempt companies that produce exams such as the Scholastic Aptitude Test (SAT), the Law School Admission Test (LSAT) and the Graduate Record Examination (GRE) are performing a public function. Yet the public has no effective recourse when an error is made or when inherent biases in the tests distort a supposedly objective measurement of intellectual ability.

Nor do consumers have any real choice if they are dissatisfied with the companies' products. There is only one SAT, only one LSAT, only one GRE. Nearly all postsecondary schools rely to some degree on test scores in making admission decisions. Too often an individual's academic future is made or broken by a test result that may not even be accurate and that may say much more about family income, race, sex and geography than about the person's skills or potential.

Many critics of our bill have greatly overstated the scope and impact of the legislation. It seeks, first of all, to give all prospective consumers uniform information about the nature of a particular exam. Just as truth-in-lending laws provide borrowers with needed data on comparative loan practices, so would truth-in-testing supply students with a helpful understanding of the meaning and limitations of the tests.

Our legislation aims to eliminate biases in the selection of questions, not by government decree, but by enabling outside researchers to review test preparation criteria and to conduct a reasonable dialogue on test methodology. This kind of independent evaluation is almost impossible at present because the testing industry treats its studies, and procedures like state secrets. Our bill guarantees access to essential material by requiring the companies to open their analyses of test results to public inspection.

Questions and answers used in computing the raw score would be released within 30 days after a test is administered, according to another provision in our bill. A test-taker would also be able to obtain a copy of his answer sheet and a list of correct answers.

This particular provision, intended to provide a cross-check against scoring errors and to show exactly where academic strengths and weaknesses lie, has been singled out for attack by the testing industry. The companies claim that this requirement will necessitate preparation of many more tests, thereby increasing cost and weakening the reliability of exam results.

Neither of these claims is valid. The legislation explicitly protects reliability by exempting from disclosure those questions and answers that do not affect the raw score and that are used by the companies for experimental purposes. Officials from the American College Testing Service testified at congressional hearings that our bill would not alter their current procedure of developing new questions for all the items in four out of five tests. The legislation also exempts from the disclosure provision exams administered to fewer than 5,000 students a year.

Truth-in-testing should have marginal impact on company expenses and on the fees charged test-takers. Only about 5 percent of the fee paid by a student taking the SAT, for example, actually goes toward test development, according to other testimony at the hearings.

Revealing questions and answers to previous tests will prove a boon to coaching schools that only well-off students can afford to attend, some critics assert. But the expensive coaching services already exist, and their clientele is almost exclusively from affluent families. Truth-in-testing is likely to make effective coaching readily available to students from all economic backgrounds since practice tests will be widely distributed.

Our legislation will improve the merit selection process for higher education by making test preparation and scoring less arbitrary. Our bill is supported by a diverse coalition of educational, minority and student organizations because it serves the interest of our educational system and of millions of students. Truth-in-testing deserves to be enacted as a workable means of replacing secrecy and insularity with openness and accountability.

Reps. Weiss (D-N.Y.), Chisholm (D-N.Y.) and Miller (D-Calif.) are cosponsors of the truth-in-testing legislation.

10-27-79 WPOST

PREPARED STATEMENT OF HON. GEORGE MILLER, A REPRESENTATIVE IN CONGRESS FROM THE STATE OF CALIFORNIA

MR. CHAIRMAN, FELLOW MEMBERS OF THE SUBCOMMITTEE, I AM PLEASED TO BE HERE TODAY AT THE OPENING OF WHAT I HOPE WILL BE AN INFORMATIVE AND CONSTRUCTIVE LOOK INTO THE WORLD OF STARDARDIZED TESTING. BOTH CONGRESSMAN GIBBONS' BILL, H.R. 3564, AND CONGRESSMAN WEISS' BILL, H.R. 4949, ARE USEFUL TOOLS FOR EXPLORING THE ROLE THAT STANDARDIZED TESTING PLAYS IN OUR EDUCATIONAL FUTURE.

IF EDUCATION DETERMINES THE FUTURE FOR AMERICANS, EDUCATIONAL TESTING LARGELY DETERMINES THAT EDUCATION. FROM IQ TESTS IN GRADE SCHOOL, TO MINIMUM COMPETENCY AND SAT TESTS IN HIGH SCHOOL, TO GRADUATE SCHOOL ADMISSIONS TESTS AND OCCUPATIONAL TESTS, WE SEEM TO BE TESTED FROM CRADLE TO GRAVE. YET THE AVERAGE AMERICAN KNOWS LITTLE ABOUT THESE TESTS, EXCEPT OF THEIR IMPORTANCE IN DETERMINING HIS OR HER FUTURE.

MANY PEOPLE, BOTH EXPERTS AND LAYPERSONS, HAVE RAISED SERIOUS QUESTIONS ABOUT THE VALIDITY AND APPROPRIATE USES OF STANDARDIZED TESTS. HOWEVER, SINCE AVAILABLE KNOWLEDGE ABOUT THEIR CONTENTS AND USES IS GENERALLY QUITE LIMITED, IT IS VERY DIFFICULT FOR EITHER A RESEARCHER OR A TEST "CONSUMER" TO MAKE AN INFORMED JUDGMENT ABOUT THESE TESTS. THIS IS REALLY

WHAT THESE BILLS ARE ALL ABOUT: KNOWLEDGE. WE ARE NEITHER DESIGNING TESTS, NOR REGULATING THEIR USE. WE ARE NOT GIVING THE FEDERAL GOVERNMENT CONTROL OVER THE CONTENT OF STANDARDIZED TESTS.

THESE BILLS DISCLOSE INFORMATION--THE INFORMATION THAT A TEST TAKER, AS A CONSUMER OF THAT PRODUCT, SHOULD HAVE A RIGHT TO KNOW. WHEN I GO TO A BANK FOR A LOAN, THE BANK MUST TELL ME ALL MY OBLIGATIONS AND LIABILITIES, AS WELL AS THE BANK'S RESPONSIBILITIES UNDER THAT LOAN AGREEMENT. WHEN I GO TO BUY A TOASTER, I GET A GUARANTEE THAT PROVIDES ME WITH CERTAIN QUALITY ASSURANCES AND A GRIEVANCE PROCEDURE IF I FEEL WRONGED. WHEN I BUY A BOX OF CORN FLAKES, I GET A FULL LISTING OF THE BOX'S CONTENTS. GIVEN THE SIGNIFICANT ROLE THAT EDUCATIONAL TESTS OCCUPY IN THE EDUCATIONAL LIVES OF AMERICANS, WE SHOULD DEMAND NO LESS FOR TEST CONSUMERS.

H.R. 4949 WILL PROVIDE THIS INFORMATION TO TEST TAKERS TO HELP THEM DETERMINE THE ROLE THAT THESE TESTS WILL FULFILL IN THEIR EDUCATIONAL FUTURE. IT WILL ALLOW STUDENTS TO KNOW HOW THE TESTS ARE REPORTED TO SCHOOLS, HOW RELIABLE THE COMPANY CONSIDERS THE TESTS, WHO WILL GET TO SEE THE STUDENTS' SCORES, HOW STUDENTS ACTUALLY PERFORMED ON CERTAIN QUESTIONS, AND HOW ACCURATELY THE TESTS ARE SCORED. THE BILL WILL MAKE MANY OF THE STUDIES THAT TEST AGENCIES HAVE DONE ON THESE TESTS AVAILABLE TO THE PUBLIC FOR THE FIRST TIME. IT WILL STRENGTHEN THE PRIVACY RIGHTS OF TEST TAKERS BY RESTRICTING DISSEMINATION

OF STUDENT TEST SCORES BEYOND INSTITUTIONS AND AGENCIES CLEARLY DESIGNATED BY THE TEST TAKER. IT WILL ALSO DISCLOSE THE COSTS OF THE TESTS. IN ADDITION, IT WILL REQUIRE THE U.S. COMMISSIONER OF EDUCATION TO REPORT TO THE CONGRESS ON THE VALIDITY AND RELATIONSHIP OF TEST SCORES TO VARIOUS ECONOMIC AND SOCIAL FACTORS.

SOME PEOPLE HAVE ASKED WHAT THE RELEASE OF ALL THIS MATERIAL WILL ACCOMPLISH. WILL WE GET IMPROVED TESTS THAT ARE MORE ACCURATE AND RELIABLE, THAT ARE FAIRER TO CERTAIN CLASSES OF TEST TAKERS, THAT PARENTS, EDUCATORS AND STUDENTS CAN TRUST? MAYBE, BUT MAYBE NOT. THAT IS UP TO THE TESTING COMPANIES AND THE AMERICAN PEOPLE, NOT THE FEDERAL GOVERNMENT, TO DECIDE. THIS LEGISLATION WILL ONLY SUPPLY THE VITAL INFORMATION NEEDED TO HAVE AN INTELLIGENT DEBATE ON THIS SUBJECT. SECRECY AND A LACK OF INFORMATION ONLY PRODUCE UNINFORMED OPINIONS AND MISTRUST OF THE TESTING COMPANIES AND THEIR TESTS. THE MORE THAT PEOPLE KNOW ABOUT TESTING, THE MORE THEY MAY CARE ABOUT THE RESULTS; AND THEY ARE BOUND TO MAKE MORE INTELLIGENT DECISIONS ON THAT BASIS.

ONE OF THE FAMILIAR SAYINGS HERE IN WASHINGTON IS, "INFORMATION IS POWER." IF THIS IS TRUE, THEN AMERICA'S TEST TAKERS ARE TRULY POWERLESS. THESE BILLS ARE DESIGNED TO GET TO THE HEART OF THIS PROBLEM: TO GIVE THE CONSUMERS OF THESE TESTS THE RIGHT TO BASIC INFORMATION SIMILAR TO THAT WE SUPPLY TO OTHER PRODUCT CONSUMERS IN THIS NATION. IT IS TIME TO GIVE TEST TAKERS THE INFORMATION, AND THE POWER, TO REGAIN CONTROL OF THEIR EDUCATIONAL LIVES.

STATEMENT OF HON. SAM GIBBONS, A REPRESENTATIVE IN CONGRESS FROM THE STATE OF FLORIDA

Mr. GIBBONS. It is a pleasure and privilege to be here today. I want to say I first became interested in this subject many years ago. As a member of the State legislature it was brought to my attention as long ago as 1953 that there was a great deal of irregularity and discrimination brought about by the testing for such common occupations as plumbing, electrical work, contractor work, and things of that sort. I thought perhaps it was something unique to my area of Florida, but I found out as I grew older and wiser that this was a problem existing in a great many places in the United States.

Let me say there is a difference in the bill I sponsored and the Weiss bill. They are not incompatible. I drafted my bill and introduced it long before I learned of the efforts that were successfully concluded in New York for which I understand the Weiss bill is patterned. If the committee so wishes these two bills could be consolidated and enacted together or parts could be selected and enacted as a bill that the committee writes.

First, let me talk about the need for this, Mr. Chairman. There are times in history when an idea's time has arrived, and certainly this is one of these. It is perhaps ironic that Hugh Carey and I sat next to each other on this committee for 6 years and worked in this Congress for some 15 years and that both of us have come to the same conclusion about the matter of testing, that people who take tests are entitled to certain rights, an opportunity to know what is in the test and an opportunity to make sure they have proper rights of appeal after the test has been given.

Every year millions of Americans are required to take tests that have significant influences on their educational and career opportunities—and this, Mr. Chairman, is where my bill differs with the Weiss bill. These tests have a substantial impact on the choice of education and employment. But those taking the tests have little information about the tests. The testing industry is unquestioned and unaccountable to consumers.

Individuals deserve to know how they are being rated and judged. It is time that the Freedom of Information laws, truth-in-lending laws, and truth-in-packaging laws be applied to the field of education and professional testing and licensing. Testing companies and professional licensing boards should be subject to greater public scrutiny. There is a need to provide individuals with minimum safeguards requiring minimum standards of due process. Actually that is what this bill is about, it is about due process.

Professional licensing tests have a substantial impact on the free flow of professional skills. Such tests almost exclusively determine entrance to over 60 occupations, including firefighters, policemen, real estate brokers, electricians, accountants, morticians, contractors, plumbers, and many more. Test results determine access to some of the most revered professions: Doctors, lawyers, foreign service officers, CIA and FBI agents. The Truth in Testing Act will not eliminate tests or lower the quality of professional services, but will add a learning element to the testing process and add an element of sunshine to professional licensing examinations. When I talk about sunshine, I really am talking about due process to those individuals involved. The bill will improve the quality of testing by making such examinations subject to scrutiny by experts and the public. It is imperative that we encourage and maintain the highest standards of quality in the delivery of professional services while minimizing undue restrictions based on secret, exclusionary testing practices.

The Truth in Testing Act basically requires that any test which is administered as a requirement for admission into any institution of higher education or any licensed profession provide detailed notice of the test subject matter and full disclosure of the examination results.

The bill has three primary objectives. First, it is an attempt to define and strengthen consumer rights. The bill requires examiners to delineate (1) what the examination is measuring, (2) a detailed breakdown of the subjects to be examined, (3) the margin of error in the test scoring, and (4) the manner in which test scores will be distributed to third parties. May I say parenthetically, here, Mr. Chairman, that many of the agencies already do this and have pretty detailed breakdowns of that.

The second purpose is to add a learning element to the testing process and to insure that the goal of the examination is educational competence and not discrimination or a subterfuge for exclusivity. This is accomplished by requiring testing boards and companies to disclose to applicants their specific performance in each subject area. Where the test is used for admission to a licensed occupation or profession, the applicant would be entitled to know what scores are needed to pass the examination. This disclosure would help the applicant assess his performance and identify subject areas needing improvement.

Third, the bill makes it unlawful to administer an educational or occupational admission test which measures knowledge or educational competence to be graded on the basis of a relative distribution of scores. In other words, let me illustrate. Suppose an examination is given for admission to a certain profession. Take architecture as an example. Suppose before the test is given or after the test is given, they decide, well, we only need 90 architects, and 100 took it, no matter how good they are we are only going to take 90. I do not say that is actually done in the architecture field, but it seems to be done in other fields.

So, this would say that the test that you took, there should be a reasonable level of competence that you should be required to have. In other words, your paper should not be graded upon a curve and just some of you thrown away. This is one of the most controversial provisions of the bill. It prohibits after-the-fact or after-the-examination curving of test scores to limit the number of admissions based on a relative distribution of scores. In other words, they test you on what you know rather than how the group performs. Obviously it is a mistake to test on just what the group knows. You may have a sorry group and many may need to be denied admission. But the test given should set a minimum standard of what is reasonable proficiency for somebody to take, and not just graded on the curve and some passed and some thrown away on some predetermined notion of how many they want to admit to that occupation or profession.

The enforcement provisions of the bill require no appppropriation and no further authorization of an additional bureaucratic agency to insure the implementation of this act. The Truth in Testing Act merely provides a test applicant with access to the Federal court system to petition for declaratory and injunctive relief to insure compliance with the provisions of the act.

Tests and other screening requirements are absolutely necessary to distinguish the abilities of individuals and to insure high levels of competency. However, the convenience of numerical labels must be balanced against the necessity to insure that examinations are used for determining a level of educational and professional competency. Certainly, no one would buy a car or a house without adequate information, but thousands of students and others must take tests each year that influence their entire future although they are denied any information about the test itself. A poor performance on a standardized test is like having an expensive car without gas—you cannot go anywhere and in many instances you do not know where to find the next gas station.

It is ironic that a country which stresses performance and achievement opts for convenient numerical test scores instead of factors such as judgment, integrity, determination, experience, idealism, and creativity, which are important factors in measuring competency and predictability.

We probably cannot reverse the heavy reliance on standardized testing but we can mandate that uniform notice and disclosure requirements exist to add a learning element to the testing process and insure due process to all individuals who are required to take a test for admission to a college or an occupation.

I urge the passage of this bill and welcome suggestions on how the bill can be improved to achieve its goal.

Thank you.

Mr. Chairman, I hope you will seriously consider this bill and take it up and pass it.

The other day, some Member of Congress sent me an item from the Detroit News. It was a story about the admission of the bar examination test in Michigan just this summer.

Apparently, someone in Michigan decided that a lot of the students should bust the exam, far higher than ever before. Some of the students were presumptive enough to protest that they were unnecessarily discriminated against.

The people who administered the test in Michigan just regraded the papers of those who complained, although there was no public notice that you had the privilege of complaining, and they passed about half of those who complained. They found in the process that even the testers had given incorrect answers to the questions.

That is one of many, many examples that takes place year in and year out around this country having to do with standardized tests. It is time we take an active interest in this.

Chairman PERKINS. Thank you for your presentation.

I see quite a bit of difference between your bill and the Weiss bill. I personally feel your bill would not be as objectionable as the Weiss proposal. I may be wrong, but the thing that has stood in the way of enacting several pieces of legislation for years and years gone by, is the old issue, State's rights. Of course in the past regulation of testing has always been left up to the States. But I am wondering whether you feel the time has now come that we should ignore that old argument. Do you feel we can get by that argument by preempting this area, as you have suggested? I would like to hear your comments along that particular line. That is the only objection I can see.

Mr. **Gibbons**. Mr. Chairman, I am not urging that the Congress or the Federal Government take over licensing or admission standards. What I am urging is that the Federal Government exercise its responsibility to make sure that there is the simple element of due process, which we guarantee to all people, the right to know what you are being tested for, and the right to know that your test paper has been properly scored, and to know how you did on that test.

I do not go as far as Mr. Weiss and as far as the State of New York went when they decided you not only had to give general information as to how you did on that test, but to say as they have said in New York, that you must give the questions and the answers. They go that far. Perhaps the committee will not want to go that far. My bill does not go that far.

So, I do not try to take over the licensing of professions or the admission of students, but I do think that the Federal Government ought to set standards and give a remedy for people who have taken tests so they will know that the test accurately reflects—or they are told what the test is supposed to cover, and once the test has been given, to make sure in the subject matter areas they know how they did in those subject matter areas.

I do not want to speak against the Weiss bill, because he has not had an opportunity to be heard, but I do realize the State of New York and our former colleague Hugh Carey, who sat on this committee for several years and was a very intelligent legislator, thought it necessary to go as far as the State of New York went.

Governor Carey signed that bill and stated his reasons for signing it. I have great respect for Hugh Carey, Congressman Weiss, and others. My bill does not go that far.

As I say, when I drafted my bill I did not even know the action in New York was pending at that time.

Chairman **Perkins**. Mr. Miller.

Mr. **Miller**. Thank you, Mr. Chairman.

Mr. Gibbons, I want to thank you for introduction of your legislation and also for your testimony this morning.

At the bottom of page 2, you say: "* * * the bill makes it unlawful to administer an educational or occupational admission test which measures knowledge or educational competence to be graded on the basis of a relative distribution of scores."

Will you explain to the committee how that would be carried out?

Mr. **Gibbons**. It just says you cannot use the old bell curve to determine how many passed and how many failed. You have to decide how much competency you are measuring and what is the minimum competency that a person must have in order to be admitted. When you grade them all on the so-called curve, you in effect are saying, well, if we have a bad lot we will take people who are less qualified than if we have a good lot from which we are selecting. That does not seem to measure competency. It seems to be a way of preventing people from admission even though they may be competent, and taking incompetents in even though they may be incompetent.

In other words the test givers should know—or someone should set in advance what is the minimum amount of competency on this test that we are attempting to measure, then those who attain this, pass; those who do not, do not pass. Those who do not pass have a right to know in which subject matter they did well in and which they did not do well in.

Mr. **Miller**. In the determination of that level of competency which is the threshold, I assume there will be some disclosure as to how that was arrived at so you do not end up with an arbitrary competency level.

Mr. **Gibbons**. If we are going to put so much reliance on tests, the test givers, the people who do this as a profession, should have some idea as to how much competency they are measuring.

Essentially now as I see it, Mr. Miller, they are not measuring competency, they are measuring competency among a particular group. If you give a test to 100 people and you say we are only taking in 90, say, for admission to practice architecture, or be a chiropractor, or to be a medical doctor, if you say we are only going to pass 90 of them, no matter how bad they are—it just does not make sense.

People who give tests ought to know what they are testing for, and they ought to be willing to say in their test, this is a passing score, not just grade it on the curve and either accept or reject those who do not happen to fit the curve, no matter what they may know.

Mr. **Miller**. Thank you.

Chairman **Perkins**. Mr. Goodling.

Mr. **Goodling**. I, too, would say to my distinguished colleague, I certainly would not wed these two pieces of legislation. I do not have nearly the problems with your legislation as I do with Mr. Weiss' legislation. I would suggest we tread lightly and see what happens in New York.

You start pushing these ideas, and soon you come to some national competency exam for promotion. That would be the most destructive thing that could happen. If you get into a national testing program you have to get into a national curriculum program in order to fit the testing program.

Mr. **Gibbons**. I recognize that, but that is one of the things my bill tries to get around. Mine is a due process bill.

Mr. **Goodling**. I suggest you do not yield to wedding.

Any testing program is questionable. Take an essay test, for example. I could always take one test and give it to six English teachers and come up with six different grades.

In section 6(c) it says: "No test shall be graded on the basis of the relative distribution of scores of other test subjects."

I am not quite sure what that means.

Mr. **Gibbons**. No occupational or admission test which tests knowledge or achievement, rather than just an aptitude test, shall be graded on the basis of relative distribution of other tests—that is a way of describing in legalese, you cannot grade them on the curve for knowledge of a subject.

In other words, the person who prepares the test ought to know the subject matter well enough to designate what is the level of competency we are attempting to measure.

They ought to know what they are testing for, and they ought to know what the minimum requirements are, and they ought to let the people know what those minimum requirements are, and not just take the good lots and the bad lots and accept people who are incompetent. It is an attempt to get around the grading on the curve which is a lazy man's way of saying well, I will take some out of this lot, and the rest will be thrown away.

That also can be used, and can be used very effectively, to exclude and can restrict entry into professions.

Mr. GOODLING. Let me specifically ask you if it means this: the current scoring system for all national tests takes the individual score and presents it as a number on the scale established by reference to an original population which took the test, and then a percentile rank. Are you saying this should not be?

Mr. GIBBONS. That should not be the determination for pass-fail. Certainly people should know how they ranked, or it is a good idea to know how you ranked, but it should not be the pass-fail mark.

Mr. GOODLING. What would be the standard?

Mr. GIBBONS. The standard for passing or failing a test would be determined by the test giver or the agency using the test.

Let us say Educational Testing Service has developed a test, I just use them as an illustration, for a specific subject matter, and an agency in my area decides they are going to use it. After some consultation with Educational Testing Service, they can determine what is the minimum score. How many out of 100 will we require a person to pass, but how many out of say 100 on a numerical scale is competency within that subject matter. If 50 percent or 60 or 70 percent of them do not measure up to that competency, they should not be passed, but if all who take the test measure up to that competency, they should be passed.

Mr. GOODLING. Then how do you determine which ones you hire or which ones you take into a class with limited space?

Mr. GIBBONS. You would have to make some kind of other determination.

Mr. GOODLING. Of course that becomes very arbitrary, then, when you do that.

Mr. GIBBONS. Well, yes. Everything we do has some degree of arbitrariness to it. My bill just says, say for admission to a profession, in a pass-fail situation—grade it on what the person knows, do not grade that person against the other people who may be in the group. If you grade against the other people who may be in the group, that becomes an easy way to limit the number of people who can practice plumbing, do electrical work, practice law, or practice medicine, or do any of those things. That is an important part of my bill. I believe that people should be able to be admitted to get out and compete in the market based upon their individual knowledge, not some arbitrary curve that denys them the opportunity to get out and compete.

Mr. Goodling. The reason I am pressing these issues is because I cannot imagine the kind of regulations we will get when this kind of legislation gets downtown.

Mr. Gibbons. Well, you bring up a good point. And I think we should be careful in our committee report or in the language of this report to detail exactly what is meant.

I want to grade people on what they know and admit them to these occupations based upon what they know, not based upon some arbitrary determination as to how many are going to pass or how many are going to fail, regardless of what they know.

Mr. Goodling. Are you also then saying you would strike down any of these special considerations that are given in order to bring about a certain amount of balance and so on?

Mr. Gibbons. You are talking about affirmative action programs and things of that sort?

Mr. Goodling. Yes.

Mr. Gibbons. Mr. Goodling, this does not go into that subject matter, although I realize that is a part of the whole process. Test them on what they know.

Mr. Goodling. One other question. In your statement you said the Truth in Testing Act basically requires any test administered as a requirement for admission into an institution of higher education or any licensed profession provide detailed notice on the test subject matter—I do not have any problems with that.

Mr. Gibbons. That is no problem. If you look at the material Educational Testing Service has furnished to the committee members, they generally do that. It is the next part they do not do. They do not tell you after you take the test how you did.

Mr. Goodling. What do you mean by full disclosure of examination results? Must they hand back the tests?

Mr. Gibbons. No. They would say in our brochure describing our tests we said we were going to test you in these different areas. Then they would tell you how you did in those areas. Suppose it was a case of—well, I am most familiar with law, so let me use a legal illustration—suppose you were taking a law exam, you did real well on the contracts part, and you did pretty well on the civil procedure part, but when you got over to criminal law you did not do worth a darn.

Well, you ought to know where you did well and where you did not do well. This would be no particular problem to them, especially on machine-graded tests, because they accumulate that information anyway, I believe. Their description of their test pretty well describes what they are giving. Maybe Educational Testing Service would do that.

I watched the "Today Show the" other day, and a representative from Educational Testing Service was on there with the National Education Association. Jane Pauley was interviewing them and thought she was going to have a scrap between them, thought Educational Testing Service and NEA would be against what was going on in New York. To her surprise and my surprise, they agreed the New York situation was fine. Well, she almost had no interview there.

But Educational Testing Service does give the test-taker a pretty good opportunity to know what that test-taker is being examined for. But the people who get the results of the examination never disclose, to my knowledge, how the test-taker did in those examinations.

Mr. GOODLING. I do not think Jane Pauley, NEA, or Educational Testing Service knows what they have done in New York. With the Weiss bill, we darn sure better wait to see what they have done in New York.

Mr. GIBBONS. As I say, I drafted my bill before I heard of the Weiss bill. They are not incompatible. What I was saying to the committee in the beginning, if you decide to take mine, you may want to incorporate some of the provisions of Weiss. If you take Weiss you may want to take some of my provisions.

Mr. GOODLING. Let me say in closing that whatever we do, I want to first of all understand what the real purpose of the legislation is; not necessarily what your real purpose is but the real purpose of any movements of this nature. I can see an awful lot of detrimental effects coming if we are not careful in what we pass and how we word it. It cannot be a pressure, for instance, for Hispanics and blacks, because we know the percentage of Hispanics and blacks entering college is increasing. I have to know whether or not this is a drive toward a national competency exam.

Mr. GIBBONS. No, sir. Frankly, this came out of my head, without me even knowing there was any kind of national constituency for it. I think the abuses that have taken place in testing have just come together and finally made a national constituency.

I think it incumbent upon this committee to try to work out a sound solution to a problem which obviously exists. Mine is essentially a due process, the same kind of due process you would give to anybody in either a civil or criminal matter as far as opportunity to know and to examine and not have either secrecy used against you or exclusiveness used against you.

Mr. GOODLING. The problem with testing is, there is too much of it. I think it is destructive, especially when I see what is done to first and second graders.

Chairman PERKINS. Tell me if I am correct, but as I look over your bill, it appears to me it would cover certification of contractors, and also real estate agents.

Is it your view that we ought to stick closely to educational testing and only in the educational area?

Mr. GIBBONS. No, sir. Mine includes both educational testing and occupational-admission testing. Some can perhaps make an argument that this bill should have gone to Interstate and Foreign Commerce or some other committee in Congress, but it was sent to your committee exclusively because frankly, I believe we live in a common market, Mr. Chairman, and we need to do the best we can to make sure it functions effectively by not having these kinds of exams that say we are just not going to let you practice in our area. No matter how bright you are we will fix this curve or exam up some way.

As I say, I first ran into it in the problem of local government when I was incorporating a city and making it larger than it had been. The people outside the old city said that is great, Mr. Gibbons, but we will never be able to do our plumbing in that area.

Chairman **Perkins**. Let me interrupt you. The second bells have rung.

Mr. Weiss, have you questions?

Mr. **Weiss**. No, Mr. Chairman.

Chairman **Perkins**. If not, let me thank you, Mr. Gibbons, for some outstanding testimony.

Mr. Weiss, you are going to testify.

Mr. **Weiss**. I am delighted to be associated in this noble venture with Mr. Gibbons.

Chairman **Perkins**. We will recess and come back and hear Mr. Weiss.

[Recess.]

Chairman **Perkins**. Do any of the members wish to ask Mr. Gibbons any questions? If you do, we will put him back on.

All right, Mr. Weiss, we are glad to welcome one of our own members here, a member of the Committee on Education and Labor.

STATEMENT OF HON. TED WEISS, A REPRESENTATIVE IN CONGRESS FROM THE STATE OF NEW YORK

Mr. **Weiss**. Thank you, Mr. Chairman.

I am going to highlight my statement and, with your permission, I would like to request that the entire statement be admitted into the record.

Mr. Chairman and members of the subcommittee, I am grateful for the opportunity to testify today on the Educational Testing Act of 1979, H.R. 4949, which I introduced July 24, along with Representatives Shirley Chisholm and George Miller.

I appreciate the chairman's decision to hold these hearings on both this legislation and H.R. 3564, the Truth in Testing Act introduced by Mr. Gibbons.

While the specific provisions of our bills differ in some respects, their overall aim is the same, to improve the accountability of the standardized testing process and to insure a greater degree of fairness in this area.

A few testing companies today exercise enormous power over the educational and occupational future of millions of Americans. The testing industry, despite its influence over vital aspects of individuals' lives, operates in a most unaccountable fashion. Like utilities, testing corporations perform a public function. Unlike utilities, they are almost totally exempt from public scrutiny.

Test results can literally change the course of a person's life. A poor score on one of these examinations often inhibits or stifles completely an individual's academic ambitions.

The tests often contain biases that result in lower scores for certain geographic, racial and income groups.

One study prepared by the University of California at Berkeley shows, for example, that test results vary according to the level of family earnings, while another study conducted by the Association of American Law Schools reveals that minorities sometimes have lower scores on the exams because of inherent biases and technical flaws in the tests.

Standardized tests have also acquired an aura of omniscience in which they are regarded by test-takers and by some academic evaluators as the single most important indication of a person's ability to perform acceptably in a particular academic setting.

All of us have heard about cutoff points below which an individual is not admitted to a school, and we are likewise familiar with the status accorded someone who has scored 700 or higher on a standardized test.

I must stress at this point that my legislation in no way attempts to dictate what questions should or should not be asked on a test, nor does it try to alter the admissions criteria developed by colleges, graduate and professional schools.

Instead, this bill attempts to make test-takers more fully informed about the standardized exam process, and it aims to lend needed emphasis to the ways in which these tests do not measure certain traits and abilities.

It is estimated that more than 50 million young Americans will take up to three standardized tests during their academic careers, while the testing industry receives at least $200 million in annual revenues.

A handful of agencies and companies develop and administer the standardized exams on behalf of a few associations or councils that indicate what and whom they want tested. The tests which these groups produce and administer include the scholastic aptitude test, SAT, the graduate record examination, GRE, the law school admission test, LSAT, the medical college admission test, MCAT, and the Miller analogies test.

More than 2,000 colleges rely to one degree or another on the SAT in assessing applicants. Nearly all medical and law schools employ the results of the MCAT and LSAT in their admission procedures.

The way in which these tests are prepared and their accuracy is not publicly known. The testing companies treat these exams like highly classified documents, withholding background information from both test-takers and reputable researchers.

The need for investigating the test development process and the grading procedure is acute.

Why is it, for example, that the Berkeley study I mentioned earlier shows that students whose family incomes are below $6,000 scored 92 points lower on the SAT than did students from families earning $30,000 or more a year, or that a Berkeley study found that high school students with high grade averages but low SAT scores did just as well in college as those students with high SAT results, or that a study conducted by David White for the National Conference of Black Lawyers indicates that on the average 120 points separate the scores of black and white students taking the LSAT.

A 1979 Federal Trade Commission report concludes that preparatory courses for the standardized exams do indeed tend to improve scores in spite of the fact that the testing companies deny that that is so. One such course is said in the FTC report to have resulted in up to a 100-point improvement in test results.

Until recently, for example, persons taking the LSAT on more than one occasion had their scores marked with an asterisk to indicate that they were repeaters. A computer was, however, at one point marking with an asterisk the scores of several persons who were taking the LSAT for the first time.

Law schools were thus supplied with inaccurate information about an individual who was asked on a separate form whether this was the first time he or she was taking the LSAT.

While many of the testing agencies enjoy tax-exempt, nonprofit status, they often produce what can only be considered a profit, that is, income exceeding expenditures.

A nonprofit, tax-exempt organization has a profit margin between 22 and 27 percent annually, and that comes out, according to some reports, to anywhere between $3 million and $3½ million a year.

The Educational Testing Act of 1979 addresses each of these concerns in five basic ways.

First, it requires that testing agencies include on the registration form for the exams a minimum amount of information about the test's purpose and methodology.

A testing agency would, under this provision, state in clear language the aim of the test, the skills being assessed, the exam's margin of error, the manner in which the score will be reported and when, the correlation of certain data between grades and test scores, on one hand, and academic and career performance on the other, the effect of preparatory courses on test scores, appeal procedure for questioning the test results or conditions, and other related information.

Most testing agencies currently provide an explanatory manual along with the registration form, but these manuals vary according to the kind and quality of information provided from test to test and agency to agency. My bill would simply insure that the information is uniform throughout.

H.R. 4949 would also require that testing agencies share with the U.S. Commissioner of Education the results of any studies or statistical analyses which they have prepared on their tests. These findings would then become part of the public record in order to allow independent study of these results.

Third, the legislation requires the testing agencies to provide the Commissioner within 30 days after a test is administered the questions used in calculating the raw score for an exam, the corresponding answers, and all criteria for translating the raw score into the score that is reported to an institution.

The test-taker would also be permitted under this provision in the bill to obtain a copy of his or her answer sheet and the correct answers upon request to the testing agency.

This provision would likewise enable independent researchers to review the questions and answers used in specific exams and thereby suggest possible improvements in test methodology,

This section of the legislation would also make the tests more relevant to a person's particular educational objectives.

It would provide it with a detailed accounting of results in each area, the test-taker will have a much better understanding of how and where to concentrate improvement efforts.

I must emphasize here, too, that the legislation requires only the release of questions and answers used in preparing the raw score, while the bill also exempts from this provision those tests administered to fewer than 5,000 subjects during a testing year.

The fourth major section of H.R. 4949 protects the privacy rights of test-takers. A student often receives no assurance that test results will not be shown to institutions or individuals to which the test-taker does not want results revealed.

Finally, the legislation stipulates that testing agencies must file information with the Commissioner of Education about their fee structures and expenditures. This provision would further insure the public accountability of the testing industry, large segments of which enjoy tax-exempt status.

I believe that the Educational Testing Act of 1979 represents a vital and overdue reform in an area that has long escaped public scrutiny and which exerts a most powerful influence over the lives of millions of our citizens.

This bill would accomplish on a national level what is already being initiated by the States. Both New York and California have, for example, already enacted truth-in-testing legislation similar to H.R. 4949. In New York, two testing groups have threatened to boycott the State unless they are exempted from provisions they do not like. Arrogance such as this only underscores the need for Federal legislation.

I hope that you agree that there is a strong and unmet need for truth in testing, a need that can be fulfilled only through enactment of legislation along the lines I have discussed today.

Thank you very much.

[Prepared statement of Representative Ted Weiss follows:]

PREPARED STATEMENT OF HON. TED WEISS, A REPRESENTATIVE IN CONGRESS FROM THE STATE OF NEW YORK

Mr. Chairman and members of the subcommittee, I am grateful for the opportunity to testify today on the Educational Testing Act of 1979 (H.R. 4949) which I introduced July 24 along with Representatives Shirley Chisholm and George Miller.

I appreciate the Chairman's decision to hold these hearings on both this legislation and H.R. 3564, the Truth in Testing Act introduced by Mr. Gibbons. While the specific provisions of our bills differ in some respects, their overall aim is the same -- to improve the accountability of the standardized testing process and to ensure a greater degree of fairness in this area.

In offering this bill, I have attempted to reform the testing procedure in a manner consistent with the basic objectives of our educational system. Standardized testing should facilitate the individual learning process and at the same time serve the broader effort to provide quality education to our people.

Under current practices, standardized tests do not meet either of these objectives.

A few testing companies today exercise enormous power over the educational and occupational future of millions of Americans. The testing industry, despite its influence over vital aspects of individuals' lives, operates in a most unaccountable fashion. Like utilities, testing corporations perform a public function. Unlike utilities, they are almost totally exempt from public scrutiny.

Test results can literally change the course of a person's life. A poor score on one of these examinations often inhibits or stifles completely an individual's academic ambitions. Banesh Hoffman, a Harvard scientist and author of The Tyranny of

Testing, argues, for example, that standardized tests "corrupt education," foster "intellectual dishonesty," penalize "depth," and reward "superficiality."

In addition, the tests often contain biases that result in lower scores for certain geographic, racial and income groups. One study prepared by the University of California at Berkeley shows, for example, that test results vary according to the level of family earnings, while another study conducted by the Association of American Law Schools reveals that minorities sometimes have lower scores on the exams because of inherent biases and technical flaws in the tests.

Standardized tests have also acquired an aura of omniscience in which they are regarded by test-takers and by some academic evaluators as the single most important indication of a person's ability to perform acceptably in a particular academic setting. All of us have heard about "cut-off points" below which an individual is not admitted to a school, and we are likewise familiar with the status accorded someone who has scored 700 or higher on a standardized test.

I must stress at this point that my legislation in no way attempts to dictate what questions should or should not be asked on a test, nor does it try to alter the admissions criteria developed by colleges, graduate and professional schools. Instead, this bill seeks to make test-takers more fully informed about the standardized exam process, and it aims to lend needed emphasis to the ways in which these tests do not measure certain traits and abilities.

To give some idea of the size, scope and power of the standardized testing industry, I would like to briefly describe some key findings by various researchers.

According to an article in the May 1, 1977 New York Times, it is estimated that more than 50 million young Americans will take up to three standardized tests during their academic careers, while the testing industry receives at least $200 million in annual revenues.

A handful of agencies and companies develop and administer the standardized exams on behalf of a few associations or councils that indicate what and whom they want tested. The agencies and companies include the American College Testing Program; the Educational Testing Service; Harcourt, Brace and Janovich and Scientific Research, a subsidiary of IBM. The associations and councils include the College Entrance Examination Board, the American Association of Medical Colleges, the American Dental Association and other professional entities. The tests which these groups produce and administer include the Scholastic Aptitude Test (SAT), the Graduate Record Examination (GRE), the Law School Admission Test (LSAT), the Medical College Admission Test (MCAT) and the Miller Analogies test.

More than 2000 colleges rely to one degree or another on the SAT in assessing applicants. Nearly all medical and law schools employ the results of the MCAT and LSAT in their admission procedures.

The way in which these tests are prepared and their accuracy are not publicly known. The testing companies treat these exams like highly classified documents, withholding background information from both test-takers and reputable researchers. The testing agencies meanwhile conduct their own studies on the validity and accuracy of their exams on a regular basis, but the findings of these analyses are not available to "outsiders."

Some agencies claim that qualified researchers can have access to some of these studies and some of the background information on test preparation. But many groups, including the California Student Lobby and the New York Public Interest Research Group, report that the companies' rules for use of these files are so restrictive as to prohibit any meaningful research.

The need for investigating the test development process and the grading procedure is acute. Why is it, for example, that the Berkeley study I mentioned earlier shows that students whose family incomes are below $6000 scored 92 points lower on the SAT than did students from families earning $30,000 or more a year?

Is it not at least a cause for additional research that a separate Berkeley study found that high school students with high grade averages but low SAT scores did just as well in college as those students with high SAT results?

And is it not most alarming that a study conducted by David White for the National Conference of Black Lawyers indicates that on the average 120 points separate the scores of black and white students taking the LSAT?

These disturbing findings demonstrate, I believe, the compelling need to open up the testing process to the student, parents, educators and the public in general.

These tests purport as well to measure general aptitude rather than specific learning techniques. Why is it then that a 1979 Federal Trade Commission report concludes that preparatory courses for the standardized exams do indeed tend to improve scores? One such course is said in the FTC report to have resulted in up to a 100-point improvement in test results.

And what about the errors or oversights that inevitably accompany processes such as these? Because of their secrecy, the testing companies are not held accountable for these mishaps despite the drastic effect they may have on an individual's career.

Until recently, for example, persons taking the LSAT on more than one occasion had their scores marked with an asterisk to indicate that they were "repeaters." A computer was, however, at one point marking with an asterisk the scores of several persons who were taking the LSAT for the first time. Law schools were thus supplied with inaccurate information about an idividual who was asked on a separate form whether this was the first time he or she was taking the LSAT.

The computer error was eventually corrected only because of an inquiry from a law school admission officer.

Another cause for public concern is the fee level developed by the testing industry. While many of the testing agencies enjoy tax-exempt, nonprofit status, they often produce what can only be considered a profit -- that is, income exceeding expenditures. The New York Public Interest Research Group reports, for example, that the Educational Testing Service -- a nonprofit, tax exempt organization, has a "profit" margin between 22 and 27 percent annually.

It would, I believe, be most interesting to students and families to learn why an $8.00 or $15.00 fee is necessary for an agency located on a 380-acre Princeton campus which includes a $3 million conference center.

The Educational Testing Act of 1979 addresses each of these concerns in five basic ways.

First, it requires that testing agencies include on the registration form for the exams a minimum amount of information about the test's purpose and methodology. A testing agency would, under this provision, state in clear language the aim of the test, the skills being assessed, the exam's margin of error, the manner in which the score will be reported and when, the correlation of certain data between grades and test scores on one hand and academic and career performance on the other, the effect of preparatory course on test scores, appeal procedure for questioning the test results or conditions and other related information.

Most testing agencies currently provide an explanatory manual along with the registration form, but these manuals vary according to the kind and quality of information provided from test to test and agency to agency. This provision in my bill would simply ensure that the information is uniform throughout.

H.R. 4949 would also require that testing agencies share with the U.S. Commissioner of Education the results of any studies or statistical analyses they prepare on their tests. These findings would then become part of the public record in order to allow independent study of these results.

Third, the legislation requires the testing agencies to provide the commissioner within 30 days after a test is administered the questions used in calculating the raw score for an exam, the corresponding answers and all criteria for translating the raw score into the score that is reported to an institution. The test-taker would also be permitted under this provision in the bill to obtain a copy of his or her answer sheet and the correct answers upon request to the testing agency.

Some of the mystery surrounding the standardized testing procedure would be ended through this provision. It would likewise enable independent researchers to review the questions and answers used in specific exams and thereby suggest possible improvements in test methodology.

This section of the legislation would also make the tests more relevant to a person's particular educational objectives. The GRE's, for example, are currently broken down into three basic categories -- verbal, quantitative and analytical. The test-taker now has only the most general knowledge of individual strengths and weaknesses under this categorization method. After taking the exam, a person will not have learned very much about his or her own abilities. But if provided with a detailed accounting of results in each area, the test-taker will have a much better understanding of how and where to concentrate improvement efforts.

I must emphasize here, too, that the legislation requires only the release of questions and answers used in preparing the raw score, while the bill also exempts from this provision those tests administered to fewer than 5000 subjects during a testing year.

This particular requirement will therefore not cause any significant increase in expenses to the testing agency and, by extension, to the test-taker.

The costs of test development are, in any event, proportionately low in comparison to revenues and other expenses. For example, the New York Public Interest Research Group reports that only 23 cents of the $7.00 fee charged for the 1971 SAT went to test question development.

The fourth major section of H.R. 4949 protects the privacy rights of test-takers. A student often receives no assurance that test results will not be shown to institutions or individuals to which the test-taker does not want results revealed.

Finally, the legislation stipulates that testing agencies must file information with the Commissioner of Education about their fee structures and expenditures. This provision would further ensure the public accountability of the testing industry, large segments of which enjoy tax-exempt status.

I believe that the Educational Testing Act of 1979 represents a vital and overdue reform in an area that has long escaped public scrutiny and which exerts a most powerful influence over the lives of millions of our citizens.

This bill would accomplish on a national level what is already being initiated by the states. Both New York and California have, for example, already enacted truth-in-testing legislation similar to H.R. 4949. In New York, two testing groups have threatened to boycott the state unless they are exempted from provisions they do not like. Arrogance such as this only underscores the need for federal legislation.

I hope that you agree that there is a strong and unmet need for truth in testing, a need that can be fulfilled only through enactment of legislation along the lines I have discussed today.

Thank you.

Chairman **Perkins**. Mr. Hawkins?

Mr. **Hawkins**. May I commend you on a very excellent analysis of the situation, a very fine presentation.

May I ask you whether or not the proposal that you are sponsoring would preempt States or whether or not States such as California would be covered?

Mr. **Weiss**. I think preempt probably is not the right word. I think what it would do would be to make the Federal standards applicable; however, any State which wanted to apply stricter standards than that which we have prescribed would still be entitled to do so.

Mr. **Hawkins**. You mentioned there is a profit made by some of the testing agencies.

In those instances, what happens to the profits? Are they distributed or in any way plowed back to the testing agencies to improve the quality of the testing or research, or just what does happen?

Mr. **Weiss**. Mr. Chairman, I don't think that any of us on the outside really know that for sure, and technically they are not profits. They would be profit margins if they were with a non-tax-exempt kind of an organization.

We do know that there has been over the course of the last 30 years or so in the case of the Educational Testing Service, the development of a magnificent center in Princeton, N.J., with a hotel and various other accommodations.

I assume in one way or another those moneys are utilized by the service, but whether they are essential utilization or just using up the money, there is no way of knowing on the basis of the information that is available.

Mr. **Hawkins**. I am not implying anything undesirable in mentioning it, but I was interested in the extent to which the testing agencies do attempt to improve the quality of their testing.

Has there been any study of how the testing that goes on by these agencies at the present time could be validated? I am not so sure the Commissioner of Education should be the one to do it necessarily, or whether the educational establishment itself should in some way set up some method of validating them, but is there any validation that goes on, any approval by any educational group other than those involved in the testing itself directly?

Mr. **Weiss**. Mr. Chairman, Mr. Hawkins, two things should be said: One, the legislation that we have submitted would not authorize any validation by the Commissioner of Education.

In essence, he would be the repositor of certain records that would be submitted by the testing services, which would then be summarized and made available to the public at large, but would not require or authorize him to do anything further by way of validation.

I don't think that there is any validation as such, except perhaps by the schools or association of schools and colleges which in essence commission the tests, but there is such an interlocking relationship that exists beween the testing service and between the people who commission the testing service to prepare the test, that it is hard to say objectively that there is independent valuation.

Some outside researchers have tried to determine whether in fact the tests are valid or not, and those are some of the examples which I was able to cite, but most independent researchers claim that they have not had access to enough information on a consistent enough basis.

Mr. **Hawkins**. As I understand the thrust of your proposal, the thrust of it seems to be in the procedural aspect of it.

I do not see it as any attempt to interfere with the actual questions that I asked; am I correct?

Mr. **Weiss**. You are absolutely right. There is no attempt to tell them what questions they ought to ask or how they ought to prepare the questions. What it simply wants is full public accountability, public exposure of the questions for the world at large, as well as for those who are taking the tests now.

Mr. **Hawkins**. Isn't the question that is asked is where the bias really prevails and, if that is so, in what way would that particular bias that is contained in the question that is asked, let's say, of a child from a low-income family being asked a type of question that would be relevant to a child of a middle class family. That disparity that is there, it seems to me, that is a larger area in which the so-called problem does exist.

Now, in what way would any of these proposals reach that particular situation, or do you feel that it should or should not be reached?

Mr. **Weiss**. It is my view and the view of those who support this kind of legislation that if you made the tests and the test results available, along with the correlation studies that have been done by the testing agencies themselves, independent scholars in the field would then be able to analyze the tests themselves and determine where, in their judgment, biases have crept into the tests themselves.

Again, it's one of those instances where if you shed enough light on it, that in itself will allow other outside objective experts to come in and tell the testing services what parts of their methods may be flawed, but this legislation would not mandate any changes at all.

Mr. **Hawkins**. Thank you very much.

Chairman **Perkins**. Mr. Goodling?

Mr. **Goodling**. Mr. Chairman, I have so many questions that it would take hours of hearing.

I hope we will have a lot of deliberation on Mr. Weiss' legislation before we move with any rapidity.

I specifically, Mr. Weiss, would like to wait to see just what New York has done, what they have created, whether they have created a monster or something good.

We will not know that for a while.

Second, the National Academy of Sciences has a committee on ability of testing which is looking into the role of standardized testing in the United States.

I would like to see their results.

Some of your provisions in my estimation, and this is a hurried judgment, of course, are going to hurt low-income individuals and minorities particularly if you start making these tests available to the prep groups. The people who are going to be able to take advantage of that are going to be those that are academically aggressive and those who have the money to pay for the coaching schools, et cetera.

So I have real concerns that the very intent of your legislation could backfire and, of course, as I said earlier, I am always concerned as to what will happen when it gets downtown.

I think Mr. Gibbons' approach has some merit. It needs a wrench to tighten it very tightly, so it cannot be misconstrued by the bureaucracy when they start promulgating regulations.

Mr. Weiss. Could I respond to the first point that you raised?

Mr. Goodling. Yes.

Mr. Weiss. You know one of the problems that exists today is that those who have the wherewithal, those who have the means to do it, attend the coaching sessions right now. Indeed, one of the open secrets in the field is that these coaching schools hire people to go in, take the tests, memorize the questions, come back and report to the coaching school what questions they remember.

They then give practice tests incorporating the coaching sessions, incorporating questions which in essence they have memorized, but the rank and file student who does not have the kind of wherewithal to allow him to pay the $120 or whatever dollars it takes for one of those coaching sessions, does not get the benefit.

What my proposal would do, I think, would be to equalize the system, so that there would be no advantage on the basis of what you could afford or what you could not afford by way of coaching sessions.

Mr. Goodling. Well, you are making it available for that purpose. But that does then give them the opportunity, because who is going to prepare them, teach them, train them?

Are they not going to have to pay for those services?

Mr. Weiss. No.

Mr. Goodling. Are they going to teach themselves?

Mr. Weiss. Again, it seems to me that once you make them public so that the tests really are available, not on the basis of how much you can afford to pay for a private coaching session or coaching group, but open to the world at large.

There are such organizations and individuals in the scholastic area which, I think, would be able to provide the kind of assistance that is needed.

Mr. Goodling. Take the SAT, for example; it doesn't test specifics.

Is there any need to have a million different SAT's released thereby forcing the testing companies to write new ones? Isn't the release of a general idea of what is on those tests and some questions from those tests enough?

Is there any need on a generalized test such as that?

Mr. Weiss. One of the questions in dispute, is precisely what does the SAT test?

Does it test scholastic aptitude?

Does it test cultural learning?

Does it test specific forms of information which are available to certain parts of the public?

Those are all open to question.

The second part that I think is not very well known, especially in light of the arguments and questions I heard in New York, is that if the New York statute stays on the books, the costs for taking those tests, and the application fees are going to be increased.

Now, because it will have to develop lots and lots of new questions, and I think many people have the impression that the same questions are given over and over again, the fact is that according to the educational testing service people themselves, some 70 percent of the questions on any particular test are new questions, in any event.

They developed new questions, and so the question really is who has access to them?

Should everybody have access, or only those people who can afford to pay the coaching schools?

Mr. GOODLING. What I was saying about SAT's was that not that the same questions were given over and over again but I am saying the same general approach to testing is the same.

I don't know whether or not there is an unlimited amount of questions that you can come up with time and time and time and time again. But you see, all of the SAT tests that have been used in the past are already available for everybody to see and everybody to use and everybody to prepare from, if they so desire.

Mr. WEISS. Well, even that is open to question as to what parts of them are open and under what circumstances.

Mr. GOODLING. Well, no, it's not open to questions. Those tests that have been used are available.

Mr. WEISS. Mr. Goodling, I wish that were absolutely true, but it is not.

In fact, we have had qualified researchers and you will hear some testimony from them, who will tell you the difficulties they have had in getting some of the test results.

Mr. GOODLING. The point I am trying to make is that I don't know how you can help an individual by releasing every test question that has been given on an SAT.

I don't know how, first of all, you could ever sit down and prepare for any examination trying to go over every one of those questions. That is an impossibility.

I am saying they are available to the general idea of what an SAT test is all about. On page 4, and certainly you can answer the question very rapidly yourself, your bill says:

> The need for investigating the test development process and the grading procedure is acute. Why is it, for example, that the Berkeley study I mentioned earlier shows that students whose family incomes are below $6,000 scored 92 points lower on the SAT than did students from families earning $30,000 or more a year?

You know all the answers to that question. What is your legislation going to do to improve that situation? What is it going to do to change the readiness of that child? What is it going to do to change the situation in the home for that child?

You know exactly why the person from that higher income in some instances and not all, scores better.

What does your legislation do in any way, shape, or form to change that home situation, that cultural situation?

What does your legislation do to make it better?

Mr. WEISS. I don't accept the premise, Mr. Goodling, that aptitude tests, not fact tests but scholastic aptitude tests, of necessity have to show that kind of a range of difference of 92 points on the basis of family income.

I think that there is something wrong about the way that those tests are constructed. You ought to be able to come up with tests which in fact tests the aptitudes of the kids taking the tests, rather than their family incomes.

There is a disagreement between us as to whether or not these tests are giving us back the obvious.

Mr. **Goodling**. What does your legislation do to improve the possibility of youngsters doing better to improve on the verbal part of the SAT?

Mr. **Weiss**. Again all that my legislation does is to open the process to the world at large, including the individuals who take the test, so that qualified experts in the field can review them, can look at them, and write and analyze and make reports as to the methodology, the preparation and the scoring of those tests.

My legislation inherently does not do anything beyond that. That is all that the legislation does.

Mr. **Goodling**. You see, and you have heard me mention this many times, I have a real fear that what your legislation could do, coupled with a new Secretary of Education and a new bureaucracy downtown, is get into the very business of standardized tests and then competency examinations coming from Washington, D.C. This would certainly be totally devastating to the system because they would have to match our testing program from on high with our curriculum.

Mr. **Miller**. Will the gentleman yield?

Mr. **Goodling**. Yes, sir.

Mr. **Miller**. How would that happen with this bill?

Mr. **Goodling**. I am trying to find out specifically what this bill is supposed to do.

Mr. **Miller**. How can you make all those remarks and read the bill and tell me how it happens?

Mr. **Goodling**. Because I have been in the business so long and I have seen what these little things that we try to do on the Federal level get translated into, once it leaves the Congress of the United States.

I say let's see what the New York system does. Let's see what the Academy of Sciences is going to come up with before we rush off into some kind of program which we are not sure will do.

I am trying to find out exactly what your legislation is supposed to do.

Mr. **Weiss**. My legislation seeks to do what I have been saying all along, that is, to simply open it up so that everybody can tell exactly what it is that they are testing and what the scores are, and so on.

The interesting thing, is that there have been misconceptions about H.R. 4949. I think that partly is because it's only a week since this bill has been in and I think Mr. Gibbons' bill has not been in too much longer. You had indicated earlier in this dialog that you thought that perhaps Mr. Gibbons' legislation provides some more safeguards but I want you to know that we drew this legislation, H.R. 4949, very, very carefully and very narrowly.

For example, there is a provision that you and Mr. Gibbons discussed as to what his legislation would do in prohibiting certain kinds of scoring.

My legislation does not get into that at all. I let the testing people and the universities and the colleges do whatever they want, prepare whatever questions they want, score them in whatever way they want. All that I am saying is that everybody ought to have the right to know what it is that they are doing.

Mr. GOODLING. Let me ask you a couple of specific things which I understand your bill is supposed to do.

Studies, evaluations, or statistical reports pertinent to a test which is prepared by a test agency, require specific names within studies, and so forth, to be protected, this is what the Commissioner of Education is supposed to do.

Specifically, what do you have in mind when you say that the Commissioner of Education to be provided the following?

Mr. WEISS. Yes; what we are asking——

Mr. GOODLING. And then what is it to do?

Mr. WEISS. We are asking that the Commissioner of Education be provided, by the testing service, the information that is set forth.

We want the Commissioner to receive and keep and hold as a matter of public record, available for people to look at, studies, evaluations, or statistical reports pertinent to a test which is prepared by a test agency.

It doesn't mandate that there be any new reports or evalutions or studies.

It says that those that are prepared ought to be submitted to the Commissioner of Education and kept there as a matter of public record so we don't have a situation where sometimes studies are completed and then they are kept, secret or not released.

What we are trying to do is to provide that the specific names of those who took tests, for example, be protected, that there be a right of privacy, that people's names ought not to be bandied about, nor what their scores were, and so forth.

A copy of contract for services between test agencies should be sent to the Commissioner of Education.

It requires that the Commissioner of Education submit a report of the materials which he has received.

That really is all that is asked for.

Mr. GOODLING. Let me press one other area quickly.

Can you tell me specifically the urgency of the bill? What is it that has come to your attention?

What do you specifically know that would say that it's very urgent to move in this direction?

What are the needs, the pressures?

Mr. WEISS. Right, right.

I think that if you had discovered that a Federal agency, the Office of Education, and the Department of HEW, were preparing and conducting tests and holding back results or releasing them in the fashion that the private testing agencies do, you would have been and would be up in arms about it. You would think that it's bureaucracy running a riot and keeping information from the public.

What has happened to us in the course of our ongoing legislative work, and especially with the publicity attendant to the work that was going on in New York, we started doing some research ourselves, and we found that going back a decade at least, there have been very, very serious questions raised about the testing process, the methodology, and the secrecy.

We decided that we were not going to get into the question of passing judgment on whether the tests are or are not valid, but we felt that we had an obligation to submit legislation which would open the process up, and that is the urgency in this matter.

It's an urgency based on information that we have developed.

Mr. **Goodling**. I am trying to get a little deeper than that. I am trying to find out what you say, is an urgency to open it up. What is the urgency to open it up?

What have been the discriminatory practices in the past?

What is it that makes it urgent other than just to say it's urgent open it up?

Mr. **Weiss**. Mr. Goodling, like you, I do not like what seems to smack of secrecy in the public's business, and I consider a situation urgent where you have some 50 million students who in the course of their academic careers are subjected to an average of three of those standardized kinds of tests and are unable to see the results of their tests.

With all of the questions that have been raised about the methodology and what they really test, I don't think that this ought to be allowed to continue. I have a built-in sense of concern about that kind of ongoing secrecy and I think that we ought to allow the breath of fresh air to go in on this situation. That is all.

Mr. **Goodling**. Can you tell me what the secrecy is in the SAT, for instance? What is the secrecy?

Mr. **Weiss**. For example, the secrecy is that a student who takes one of those tests does not understand the present system of the right to request of the people who prepare and administer that test, that his paper with the raw scores be given back to him or her, so that in fact the test taker knows how they did on the test.

Why should that be?

Mr. **Goodling**. Well, simply because, as you know, if you are going to release every one of those tests, then you are going to have to prepare a new one over and over again, and I don't know whether you can keep coming up with good, logical questions on an SAT.

Mr. **Weiss**. They give those tests sometimes six or nine times a year; 70 percent of the tests on any examination are new. The question really is: Should it be a closely-held secret or should it be open to the world?

That is really the question.

Chairman **Perkins**. Let me say to my distinguished colleague from Pennsylvania, I don't want to cut him off but we are not going to get through today if we go ahead with questions now.

Mr. **Goodling**. We should take a lot of time and see what to do. I won't take any more time.

Chairman **Perkins**. The gentleman almost ran 20 minutes.

Go ahead, Mr. Miller.

Mr. **Miller**. I won't take much time.

I just want to tell Congressman Weiss that I think he is right, and I think you have a process that makes a number of determinations about how people will spend the rest of their lives.

We heard earlier from Mr. Gibbons about whether or not people would be allowed to engage in various occupations, and now you are going into the question of whether they will be allowed to enter various colleges, various professional schools to pursue other occupations.

I think that the point is that when you have an entity that can make those kinds of determinations, even assuming that all of the determinations are valid and can be supported through analysis of the procedures, the people who participate and are required to participate—this is not something that is optional—ought to know all there is to know about the process.

You don't tell the University of California you don't feel like taking the SAT's. You don't tell the law school that you don't want to take the LSAT because you feel you are competent to go to law school.

You have to take it, and in that case the people who engage that process as test-takers ought to know all there is to know about the process that is going to guide their lives, and they ought to have the ability to challenge it and review their tests and they ought to be able to sit down and go over it to see what is going on.

They ought to be able to understand the manner in which those scores are developed, the biases that are built in or are not built in, and I think that it's very important for this kind of information.

To go from there and suggest a national curriculum, to suggest national testing as the result of a disclosure bill, I think is simply to set up strawmen so that they can always be out there to bugaboo the bill, and I don't think that is the situation at all.

First of all, this committee would never go for it, and second, it's not related to this issue.

The issue is disclosure to young people who have to make academic determinations and who want to know how those determinations are going to be made for them.

I would be delighted if the author of this legislation would point out where it gives the basis to talk about national testing, national test development or national curriculum.

I just don't see it in this legislation. We ought to address ourselves to this legislation, and I don't see it in the bill, Ted, and I want to give you a chance to respond to that, because it's one thing to talk about it but it's not in the legislation, as I read it.

Mr. WEISS. You are absolutely right, Mr. Miller.

It's not in the legislation.

The legislation was very, very carefully drawn.

There is no suggestion of any national tests, and there is nothing in there even impacting on the use of these tests made by the colleges or universities or the professional schools.

All of that really is left exactly as it is. All that we are saying is that we ought to really open it up so that people see it.

I should also remind my good friend, Mr. Goodling, I share some of his concerns about national curriculae and national departments of education, and so on, and I opposed the creation of the Department of Education and have been voting it in the Committee on

Government Operations and on the floor of the House for the better part of 4 years now.

I think that the danger, sometimes when we look at labels, is that we allow people to get away with doing things in a fashion which we would never in a million years allow a Government agency to do. The testing companies in this instance, I think, are in that position.

They have assumed, they could continue, because nobody has challenged them to do things in a fashion which would never condone the Government agency.

Mr. GOODLING. Would the gentleman yield?

Mr. MILLER. Sure.

Mr. GOODLING. You should certainly include the civil service in your legislation, if you are going to make a statement like you just made when you talk about what we will allow and will not allow, when it comes to Government agencies and Government-connected agencies, and so on. I think you will find that if ever there is a place that needs your proposal it's probably civil service testing.

Mr. WEISS. My legislation?

Mr. GOODLING. You might think about that.

Mr. WEISS. My legislation doesn't even address the area that Mr. Gibbons' legislation addresses, that is, the occupational testing.

All that this legislation does, because I think your admonition in some regards is a very wise one, is to start very slowly.

It talks about admissions to colleges, the professional schools, and the graduate schools period. That is it.

Mr. GOODLING. I was just responding to your comment that we wouldn't allow such a thing to be connected with the Federal Government, and we do.

Chairman PERKINS. We have a vote on the Floor.

We will recess for 10 minutes and resume the hearing.

[A brief recess was taken.]

Chairman PERKINS. The subcommittee will be in order.

We are waiting for Mr. Weiss to come back. He is voting. We will wait a couple of minutes.

Why doesn't the next panel come around?

Dr. Walter Haney, testing expert; Dr. John Cooper, president, Association of American Medical Colleges; Mr. Fred Hargadon, chairman, board of trustees, the College Board, accompanied by Ms. Lois D. Rice, vice president, Mr. Robert J. Solomon, executive vice president and acting president, Educational Testing Service.

We will hear from you first, Mr. Hargadon.

STATEMENTS OF DR. WALTER HANEY, TESTING EXPERT, ACCOMPANIED BY ALLAN NAIRN; DR. JOHN A. D. COOPER, PRESIDENT, ASSOCIATION OF AMERICAN MEDICAL COLLEGES; FRED HARGADON, CHAIRMAN, BOARD OF TRUSTEES, THE COLLEGE BOARD, ACCOMPANIED BY LOIS D. RICE, VICE PRESIDENT; ROBERT J. SOLOMON, EXECUTIVE VICE PRESIDENT, ACTING PRESIDENT, EDUCATIONAL TESTING SERVICE

STATEMENT OF LOIS D. RICE, VICE PRESIDENT, COLLEGE BOARD

Ms. Rice. Mr. Chairman, my name is Lois D. Rice, and I am a vice president of the College Board.

Unlike many other occasions when I have appeared before you and your Subcommittee on Postsecondary Education, today I am not the witness.

Instead I have the pleasure of introducing the witness for the College Board, an organization which has had the pleasure, Mr. Chairman, of supporting your efforts to bring about a more just society by breaking down the barriers to educational opportunity.

We walked together on a torturous road in 1972, and without, as you know, much consensus in the higher education community to enact what is now the centerpiece of Federal support for postsecondary education—the basic educational opportunity grants program.

Chairman Perkins. May I make an observation that you were very helpful in that long and tedious and complicated fight, Ms. Rice, and I certainly appreciate it.

Ms. Rice. I thank you, Mr. Chairman.

Presenting our testimony is Fred Hargadon, a political scientist, a dean of admissions and now the elected chairman of the College Board.

Mr. Hargadon has always acted on his beliefs. He has devoted himself to educational services for the disadvantaged—to upward bound and talent search programs at the local, State, and national levels.

I am, therefore, pleased and proud to introduce him to the committee. He is perhaps the Nation's most quoted admissions' officer in the press, but one who has not until today had the honor of speaking before a congressional committee.

STATEMENT OF FRED HARGADON, CHAIRMAN, BOARD OF TRUSTEES, THE COLLEGE BOARD

Mr. Hargadon. Mr. Chairman and members of the subcommittee, my name is Fred Hargadon.

For the last 10 years I have been dean of admissions at Stanford University and prior to that dean of admissions for 4 years at Swarthmore College.

I also currently serve as the elected chairman of the College Board.

I welcome this opportunity to testify on H.R. 3564, Representative Gibbons' Truth in Testing Act, and on Representative Weiss' Educational Testing Act of 1979, H.R. 4949.

Before speaking directly to the specifics of these bills, I would make two general observations.

First, I think these bills, if enacted into legislation, would significantly alter the historic role of the Federal Government in education. They might actually impede the goal of equalizing educational opportunity, a national goal, and the historic, unchanging goal for which the College Board was founded.

Second, both bills, though in different degree and through different means, would put the Federal Government into the business of regulating testing programs and admissions practices in ways which run the risk of increasing the cost to students, educational institutions, and society as a whole.

Tests do affect the lives of students, although nowhere near as significantly or as negatively as the proponents of this legislation would have us believe. In fact, when it comes to college admissions, it is the quality of students' prior educational preparation and the quality of their performance, as indicated by the grades they have received, which play a greater role than that of their results on College Board tests.

Despite this fact, it is tests to which some legislators at the State and Federal level have recently turned their attention.

We sympathize with the quest for accurate and full disclosure of information to those who take our tests. At the same time, we are concerned about the hazards of precipitous legislative action and the unanticipated, unintended effects that often result from the Government intervention into educational affairs.

The Buckley amendment is a prime example.

We sympathize with the stated intent of the framers of that measure, the Family Educational Rights and Privacy Act, FERPA, of 1975. But the legislation touched off an agonizing and time-consuming ordeal of statutory interpretation and regulation writing.

And despite the passion that inspired the legislation, few parents and students seem to know or care about the procedures which allow them access to information in student files.

The legislation now before this subcommittee could generate the same kind of cumbersome, costly, and sweeping regulation, and without any clear documentation that the changes it seeks are either needed or useful.

What is the present situation in undergraduate admissions testing? The College Board is responsible for testing more students for undergraduate college admission than all other organizations combined, more than 1.5 million in 1978-79. We are a nonprofit, voluntary association of more than 2,500 schools and school systems, colleges and universities and educational associations. Our two main activities are the admissions testing program, the most familiar part of which is the Scholastic Aptitude Test—the SAT—and the College Scholarship Service—CSS—which last year served 2.6 million aid applicants.

Since the founding of the College Board more than 75 years ago, our primary commitment has been to promote, not to retard, access to higher education in our society. In the years following World War II, the admissions testing program enabled a vast expansion of postsecondary education by providing colleges and universities with an additional instrument for measuring academic potential among our young people. In the 1950's the College Board membership responded to the need for a more equitable distribution of financial aid by pioneering procedures for awarding such aid according to financial need, a move that increased educational opportunities for the less affluent and raised the level of participation in postsecondary education of minority students. At the time of sputnik, with the Nation searching for higher standards of excellence for its young people, the College Board initiated the advanced placement program—a model for providing excellence in the secondary schools and a program in which over 100,000 high school students participated this year.

And the College Board has developed and sustained these efforts without prodding from any government.

Also without prodding from any government, we provide those who take our tests with nearly all the information about tests which the bills before you today would mandate.

Consistent with section 6 of H.R. 3564 and section 3 of H.R. 4949, the College Board now provides detailed information to students describing the tests and their purposes, what they do and do not measure, their reliability, their predictive validity, the student's performance relative to other students, how scores are calculated, and when they will be reported to the student and institutions. We supply all this information and more about the SAT and achievement tests in pamphlets widely available to high school students. Every student registering for the scholastic aptitude test automatically receives the publication "Taking the SAT," which includes for practice purposes a complete sample of a real test, along with an answer key, explanations of questions and answers, and instructions for timing and scoring the test. And following each test administration, every student also receives a complete and careful explanation of his actual score, which is sent only to institutions designated by the test taker. High schools and colleges all receive periodic College Board publications explaining the SAT and the technical limits on the uses of SAT scores.

Mr. Chairman, I wish to give the committee for the record a complete set of materials the College Board voluntarily provides to students. But let me tell you how students respond to them.

[Information referred to above is retained in subcommittee files.]

Mr. HARGADON. In a 1977 sample survey of 6,000-SAT takers, conducted by the Response Analysis Corp., we found that 48 percent of students read all of the pretest information, and 60 percent tried to answer all or most of the practice test. Only 32 percent read all or most of the post test interpretive materials—which suggests that if we want students to approach the tests as a learning exercise—one of the stated aims of the legislation's proponents—opportunities for self-assessment beforehand are likely to be more effective than post test disclosure and review, as called for in H.R. 4949.

The sole purpose of the admissions testing program and the SAT is to provide a supplemental, but common measure of ability for secondary school students who have taken different courses in different high schools with different grading practices.

The waste of resources has been significantly minimized at most colleges in this country. It used to be colleges would take anybody and plan to flunk out half of their students by the summer of their first year. This does not happen when tests are used as a supplemental tool in the admissions process.

Over the years there have been charges that the SAT is biased against this group or that group, with a low score being the evidence of the bias. We of the College Board have of course been sensitive to such charges, for it would be a cruel anomaly if an association of educational institutions devoted to increasing access to college were in fact putting up, instead of knocking down, barriers to higher education.

And so we have sponsored almost continuous research devoted to detecting and rooting out any bias in our tests, and every responsible evaluation of those tests for 20 years now shows we have succeeded. Remember that the SAT's function is solely to help predict academic performance in the first year of college. So it should come as no surprise that, on average, students from families where neither parent is a high school graduate, and where there are no books in the house, and who receive poor preparation in grade school and in high school, will not do as well on the SAT as test takers whose parents were both college graduates, and the father has never been unemployed, and who have received first-rate preparation from kindergarten through the 12th grade. The SAT takes society as it is, not as it ought to be. If disadvantaged groups tend to achieve lower scores, and they do, the solution is not to shoot the test administrator, but to do something about the causes of disadvantage, and that is exactly what your committee, Mr. Chairman, has been attempting to do through the Elementary and Secondary and Higher Education Acts.

Turning now to the specific subjects of today's hearing, I want first to discuss H.R. 3564, the bill sponsored by Mr. Gibbons. It prohibits the giving of any educational or occupational admissions test unless prescribed information is provided the test taker beforehand and unless other prescribed information is furnished on request afterward; forbids grading such tests "on the basis of the relative distribution of scores of other test subjects"; and establishes a right to seek Federal judicial relief for any actual or threatened violation of the proposed act.

While we are not sure of the technical meaning to be given all the terms in H.R. 3564, to the extent that our interpretations are correct, the College Board already provides its test takers with all the information called for by the bill, with one large exception.

That exception is the bill's requirement that every test taker be told the "score which is generally required for admission to institutions of higher education." There is no such score. It is simply—and fortunately—impossible to generalize about the diversity of admissions standards and procedures followed by hundreds of colleges and universities using results from our admissions testing program. More importantly, the College Board discourages the use of score cutoffs or the use of scores in isolation from a variety of other considerations in evaluating candidates for admission. In fact, one of our major concerns with the Gibbons bill is that this provision might focus undue attention on objective criteria and, therefore, have a chilling effect on national efforts to diversify student bodies.

In sum, with the exception I have noted, the passage of H.R. 3564 would have no effect on the College Board and the 1.5 million young people who annually take our tests. But I do not want to suggest that we are indifferent to its enactment.

First, the Gibbons bill also affects every test used "as part or all of the basis for admitting or denying admission to an individual to any occupation in or affecting interstate commerce," and defines test so broadly that it necessarily includes all kinds of State licensing examinations, from doctor to lawyer, from plumber to cosmetologist, from real estate salesman to architect, and so on. We are unaware that a case has been made, much less offered, for Federal regulation of occupational testing on the scale contemplated by the Gibbons bill.

Second, even though H.R. 3564 in its workable aspects only tells the College Board to do what it is already doing, in our view harmlessness is not a reason for passing a law. A bill that does not do anything is not worth passing, but if it does get passed, it becomes the predicate for later legislative proposals that will indeed increase Federal control over the doings of our citizenry.

We also urge you not to act favorably on H.R. 4949, the Educational Testing Act of 1979, sponsored by Mr. Weiss, Mrs. Chisholm, and Mr. Miller. H.R. 4949 is far more intrusive than the Gibbons bill, for it not only directs the College Board to do what the Board is already doing, but it orders the release of all standardized tests shortly after the first time they are administered, and it puts the U.S. Commissioner of Education into the business of regulating and evaluating college admissions testing programs.

The first substantive part of the Weiss bill is section 3. That section prescribes in considerable detail just what a testing service must tell test takers ahead of time. By any reasonable standard, our pamphlet, Taking the SAT, which goes to all test takers, meets the requirements of section 3. That section demands some information that does not exist, for example, the average score increase obtainable by attending a coaching school. I think Mr. Weiss in his earlier remarks did not accurately quote the full FTC coaching study. We still believe coaching has no significant effect—but technical problems of that sort could be corrected if the bill otherwise warranted passage. So our serious concern about H.R. 4949 does not stem from its section 3.

Our main concerns begin with section 4. That section requires furnishing the U.S. Commissioner of Education with research reports on our tests and with copies of any service contracts we may have with other testing organizations. The Commissioner is given 1 year to prepare a report to Congress on the relation between test scores and income, race, sex, ethnicity, and handicapped status.

Obviously we can have no objection to the dissemination of research results in connection with our testing programs, and our reports are already available to the Office of Education. But we do not understand why the Congress should order us to furnish the Commissioner with a copy, for example, of our contract with the Educational Testing Service, any more than the Congress should ask colleges and universities to file with the Commissioner their many contracts with outside agencies. If the College Board and other testing organizations are to be singled out in that fashion, there should be a clearly stated public purpose for doing so; and there has been none stated in this instance.

Section 5 requires furnishing the U.S. Commissioner of Education with copies of all standardized tests, plus the correct answers and other data, within 30 days after scores on such tests have been released to test takers. The material is specifically to be made available for public copying. Additionally, any test taker is to be entitled, on request, to receive a copy of the test he took plus his and the correct answers. He may be charged only a nominal fee for such service, defined in the bill as "not to exceed the marginal cost of providing the information." The requirement of filing tests with the Commissioner, and for sending tests and answers to test takers, does not apply if the anticipated number of annual test takers for a given test is less than 5,000.

If section 5 were to become law, the various tests sponsored by the College Board could be administered only once before their security was breached and their validity destroyed.

At present we make the SAT available more than 20 times each year, not counting the dozens of individual one-on-one administrations for handicapped students who must choose their own place and their own time to take the test. If we are to make public the questions used on these tests each time, then the cost of developing tests each year would necessarily increase. Test disclosure will increase the need for and, therefore, the cost of test development.

Or if the increased test development costs are to be restrained, the tradeoff is to reduce the opportunities for test taking by the student population. I would like every student to be able to take home his test form; but if the test questions are to be made available, every time, to those who take the tests, then the cost will go up, and the opportunities for taking the test will go down.

The central question, then, is whether that tradeoff provides any substantial gain from the standpoint of student interests. We believe it does not.

It has been the policy of the College Board, throughout its history, to make the SAT available as widely as possible among the country's high school population at the cheapest possible cost and with the maximum flexibility to accommodate the special needs of students. This legislation proposes a different principle and a different standard. It suggests that students should be prepared to pay more or to have tests less available in order to secure the opportunity of reviewing the test upon receiving their scores.

I know of no educational argument for this, and it is not at all clear on what basis the proponents of this legislation reached this conclusion. Publication of test questions gives them undue emphasis. Teachers may be asked, then, to "teach to the tests," and this legislation will, in turn, have begun to influence the curriculum.

We are not certain at this point whether the test researchers, technicians, and scholars who now jointly develop and validate our tests would be able to multiply the number of test versions while still preserving their current high quality in terms of reliability and validity. But we are certain that research and development costs would have to rise, and that the higher costs would inevitably be reflected in higher fees by all test takers.

Given the research uncertainties and the cost certainties, Congress should look long and hard at whatever arguments are advanced in favor of enacting section 5. It is our strong view that the issue is not yet ripe for a legislative decision.

We are also troubled to some extent by the constitutional issue lurking in section 5. After all, our tests do constitute a valuable copyrighted property with a predictable useful life. The filing requirement of section 5 would substantially shorten that life, but the bill neither authorizes nor provides any compensation for the property it would take away.

Our final observation about section 5 has to do with the provision directing that we may charge only a "nominal fee" for providing additional information to test takers. We are, as I have noted before, a nonprofit association. As a nonprofit association, we publish full financial statements each year, and further detail is freely available from our treasurer. The College Board has neither endowments nor property; its revenues and expenditures have fallen within 2 percent of each other over the past 10 years.

We have throughout our history kept our service charges at the lowest possible levels and we will continue to do so. For example, in the academic year 1978-79 the charge for the SAT was $8.25—up from $7.25 20 years ago, but less than half as much in real terms, and the fee is waived entirely for those who cannot afford to pay it, and this was done for 32,000 test takers last year including all in Upward Bound programs.

We do not believe there is any reason for the Congress to pass a law regulating the charges we may make, or mandating that some users of College Board services will pay less than a fair share of the costs of those services.

Section 6 prohibits release of test scores unless specifically authorized by the test taker. The Buckley amendment enacted in 1974 contains a similar provision, but with an exception permitting carefully guarded test score disclosures for test validation purposes. We suggest that the Congress does not have to legislate again on the same subject.

According to section 7, and "In order to insure that tests are being offered at a reasonable cost," all testing organizations would have to file detailed revenue and expense information with the U.S. Commissioner every year. Presumably, the Commissioner is to make a judgment as to whether the charges to test takers are reasonable.

Finally, the bill directs the U.S. Commissioner to issue implementing regulations within 120 days of enactment, and provides a $2,000 fine for any violation on the part of a testing organization.

As you know, the Gibbons bill and the Weiss bill are not the only legislative proposals with respect to the college admissions testing process. California and New York have recently legislated in the same field, and other State legislatures have been asked to do so. Most of the publicity on the subject has focused on the question of whether a test taker should be entitled as a matter of right to obtain a copy of the test he took. There are arguments on both sides of that issue, and one of the two main points I want to make with you today is that the debate is not ready to be settled. We want to support the idea of test disclosure, but we want also to be sure that we can maintain a high quality, reliable, and valid admissions testing program at the least possible cost to those who take our tests. And we want to maintain a level of services that insures convenience to all test takers.

There is not a public emergency in connection with our testing program and no need to legislate hurriedly on the basis of slogans, misunderstandings, and incomplete or inaccurate facts. A good reminder when the term "public interest" is used, is that there are in fact many public interests. There is a need to be deliberate and prudent in legislating change of the magnitude contemplated by H.R. 4949. Consequently, we urge that you let this issue ferment for a while yet, so that we can all see and learn from the experience gained in connection with the State legislation already enacted, and so the College Board and other testing organizations can have time to explore more thoroughly the feasibility and cost of a testing program that would satisfy both the proponents of test disclosure and the proponents of a testing program that produces reliable and valid results.

The other main point I want to make is that Federal legislation with respect to higher education is a particularly sensitive and delicate matter. Over the years the Congress in general, Mr. Chairman, and your committee, in particular, have been generous to higher education in the United States. But you have also been quite consistent in keeping the Federal Government out of the business of deciding what shall be taught, how it shall be taught, who shall teach it, and who shall be taught. The bills we have been talking about this morning are the start of Federal regulation of college admissions testing programs, just one step removed from Federal regulation of college admissions themselves. We at the College Board do not believe you have thus far been given a basis for starting down that road.

I thank you, Mr. Chairman. I shall be glad to answer questions.

Chairman **PERKINS**. Without objection, we will defer questions until we hear from all members of the panel.

The next witness is Dr. Walter Haney, testing expert.

Go ahead, Dr. Haney.

STATEMENT OF DR. WALTER HANEY, TESTING EXPERT, ACCOMPANIED BY ALLAN NAIRN

Dr. **HANEY**. My name is Dr. Walter Haney. I am staff director of the National Consortium on Testing and project director at the Huron Institute of Cambridge, Mass. I am pleased to have the opportunity to come before this committee to offer testimony concerning the legislation embodied in the proposed Truth in Testing Act of 1979 (H.R. 3564) sponsored by Congressman Gibbons, and the proposed Educational Testing Act of 1979 (H.R. 4949) sponsored by Congressmen Weiss and Miller, and Congresswoman Chisholm. Let me make clear that my comments this morning represent my own professional and personal opinion, and do not necessarily reflect those of the institutions with which I am affiliated. Nevertheless, I would like to point out that the essential goals behind these proposals—namely, to increase public understanding of the strengths and weaknesses of standardized testing and assessment and to improve testing and assessment practices—are ones which prompted the founding of the National Consortium on Testing in 1977.

The role of standardized testing in American education and in American society generally is presently both highly prominent and increasingly precarious. Testing has clearly been a growth industry over the last two decades. It has been tied, for example, to increasing public concern over accountability in education and to mushrooming government enthusiasm for program evaluation. At the same time, however, there has been increasing concern over the shortcomings and ill effects of standardized testing. The technical rationale underlying standardized tests is subject to increasingly critical scrutiny. There is growing worry over the distorting effects of testing on educational goals and practices. More and more people are becoming alerted to the misuse of test results in labeling children in unfair and discriminatory manner. The two legislative proposals before this committee clearly reflect this increasing public concern over standardized testing in our society.

Before commenting on the specific provisions of the proposed legislation I would like to interject a note of caution. H.R. 4949 would require reporting of a wide range of information on tests and test use to the Federal Government. Some of these proposed requirements strike me as somewhat ironic since it is fairly clear that substantial misuse of standardized tests over the last decade has not simply grown out of the practices or policies of local educational institutions or test publishers but has been prompted in significant measure by reporting requirements of Federal and State governments.

Having expressed that general reservation, let me focus in the remainder of my remarks on what I think may be the most important aspect of the proposed legislation—namely, what might be called the sunshine provisions of H.R. 4949—section 5—which would require public disclosure of test contents, and upon request, return of corrected test questions to individual test takers.

As I am sure you are aware, proposals for so-called open testing have emerged recently in legislative initiatives in several States, most prominently in legislation passed in New York this year. Several arguments have been advanced in favor of, and in opposition to, such open testing proposals. Since there is not time today to review each of these several arguments, let me focus on what I see as the main issue: The primary argument in favor of open testing is the proposition that as a matter of simple justice, individual test takers have a right to know the exact manner in which they are judged, and the details of the assessments which may affect their educational opportunities and life chances. The main arguments in opposition to open testing, which as far as I can tell originate mainly from publishers and sponsors of currently secure tests, are that open testing would disrupt current procedures for developing, pretesting, and equating tests. In part these arguments—that open testing should be opposed because it would disrupt secure testing procedures—strike me as quite ironic. It is the obvious intention of the open testing proposals to do just that—namely, to disrupt current secure testing procedures.

Nevertheless, let me deal with specifics of the arguments against open testing. Arguments dealing with pretesting and equating of tests are essentially irrelevant to H.R. 4949 since the bill's current provisions would, as I understand it, allow for maintaining security with respect to items which are used solely for the purposes of pretesting and equating, and which do not contribute directly to an individual test-taker's raw scores. The main argument against open testing, apart from possible cost issues, seems then to be that open testing would diminish the validity, reliability, and general technical quality of tests which would under open testing have to be developed essentially anew for each test administration.

One critic of open testing, for example, has asserted that "To use materials from active item pools * * * would invalidate them for future reuse."[1] Such flat assertions of item invalidation as necessarily resulting from open testing are simply erroneous. This is clear from the fact that several sorts of test questions are currently publicly available and still can be validly used—essay test questions, and publicly released items from National Assessment now being used in several State assessment programs, are just two examples.

As a result, it is not appropriate to view the issue of test validation as an all-or-nothing issue, but rather to think in terms of the conditions under which, and the degree to which, public release would result in diminished validity. I can, for example, mention three possible mechanisms for developing tests which would allow for release of corrected tests after administration, and would at the same time minimize threats to test validity:

First, reliance on test items which directly measure relevant skills, essay questions for example, rather than indirect indicators of real skills as is the case with multiple choice items;

Second, development of item generating algorithms, which could produce large numbers of items equivalent in terms of content;

And this has been proposed recently by several people, and that is:

Third, development of large test item banks from which relatively small samples of items could systematically be drawn so as to yield alternative test forms. Essentially all I am saying here is there are certain techniques that would be used which would minimize the concerns about test validity.

A second perspective to which I would like to draw attention concerns the current quality of secure tests.

Publishers and sponsors of such tests naturally now maintain that the procedures for developing and reviewing their instruments help guarantee high quality.

The Educational Testing Service, for example, has asserted that actual tests for virtually all national testing programs are available for professional scrutiny and have been for many years.

The Mental Measurements Yearbook is probably the most famous example of such external professional review, which typically includes scrutiny of test questions against specifications and statistical reports.[2]

---

[1] Carmen Findley and Frances Berdie. "The National Assessment Approach to Exercise Development," Washington, D.C.: National Center for Educational Research and Development, Office of Education, 1970, p. 5.

But I would point out that in a speech delivered about 2 years ago at the University of Iowa in March 1977, Oscar K. Buros, educator of the Mental Measurements Yearbook, put forth a quite contrary view.

He declared:

> The information available to permit an adequate assessment to be made of these secure tests is quite unsatisfactory. Although our reviewers generally receive some in-house material, which is not available to other educators and psychologists, even this material is inadequate.[3]

In short, the assertion by proponents of secure testing that current procedures for review and disclosure of information on such tests are sufficient to guarantee high quality test instruments is challenged not only by open testing advocates but also by the very author whose work has been cited to bolster the assertion.

In closing, let me note that I have not addressed several pertinent issues, dealing with costs of open testing, test bias and test coaching, for example.

There is a great deal that could be said about these, but there is really not time for me to communicate my thoughts on some of these issues.

I would, of course, be happy to answer questions on these issues.

My own summary view is that if one accepts the validity of the proposition that individual test takers have a right to know the basis on which they are judged in matters that will affect their educational and life chances in significant ways, and I would like to make clear that personally I do, then issues of test development and technical quality need not stand in the way of open testing policies designed to implement that right.

Thank you.

Mr. WEISS. Thank you very much, Dr. Haney.

The next witness will be Dr. John Cooper, president, Association of American Medical Colleges.

STATEMENT OF DR. JOHN A. D. COOPER, PRESIDENT, ASSOCIATION OF AMERICAN MEDICAL COLLEGES

Dr. COOPER. Thank you very much, Mr. Weiss.

As you know, the association is a nonprofit educational organization whose membership includes all 125 medical schools in the United States, their teaching hospitals, representatives of their student bodies and faculty organizations, including all medical college admission deans.

The association sponsors the new medical college admission test, new MCAT, on behalf of and under the surveillance of its member schools. Canadian medical schools and American osteopathic medical schools have elected to also make use of the test.

Mr. Chairman, the legislation you are considering today is unwittingly consistent with a nationwide effort by some to eliminate or emasculate standardized testing. This philosophy is sapping the quality of education in this country at all levels.

---

[2] Educational Testing Service. "Memorandum of Opposition [to the 1978 Lavalle bill]." Undated [1978].

[3] Oscar Buros. "Fifty Years in Testing: Some Reminiscences, Criticisms and Suggestions," *Educational Researcher*, July-August 1977, p. 14.

It results in inflated grades and leaves many high school graduates functionally illiterate. The strength of this country will not be maintained by marching to their deceptive banner, truth in testing. We should embrace the more worthy concept of truth in education which standardized testing fosters.

We cannot speak for the effects of the two bills on general standardized tests such as college boards and American college testing (ACT) assessment. However, as Mr. Perkins surmised, we can state that your bill, H.R. 4949, the Educational Testing Act of 1979, we believe, is more intrusive and poses a serious threat to highly specialized tests, such as the new MCAT.

It makes allegations about standardized testing that are not consistent with the situation for the new MCAT. It indicts the new MCAT on the basis of suspicions about very different tests.

The association has no quarrel with the stated purposes of the bill, but objects seriously to the way in which these goals are used to justify what, in our view, is unnecessary and counterproductive regulation of highly differentiated tests such as the new MCAT.

Indeed, section 2(b) of your bill could well have been derived from the test objectives and guidelines for the development and implementation of the new MCAT examinations.

In my testimony I intend to describe:

How we insure test subjects and those who utilize test results are made fully aware of the characteristics, proper uses and limitations of the test (purpose 1 in section 2(b));

How we make available extensive public information regarding the procedures for development and administration of the new MCAT (purpose 2);

How the public interest is protected through promotion of knowledge about the proper use of the new MCAT and the multiple steps we take to assure accuracy, validity, and reliability in the development, administration, and interpretation of the test results (purpose 3).

The committee is undoubtedly aware of the association's decision to discontinue administration of the new MCAT examination in the State of New York after January 1980 under the requirements of the recently enacted law in the State.

Our opposition to this bill comes only after long and careful consideration of its impact. In our judgment, the impact of the New York law and these bills as currently structured seriously affect our ability to maintain the integrity, high quality, and usefulness of our examination as one criterion in the highly complex process for evaluation of medical school applicants.

This admissions process must assure equity for the applicant, but this is not all. It must select candidates capable of attaining the high level of scientific understanding, human compassion, and clinical performance that makes them worthy of the trust and confidence that all citizens have a right to expect in their personal physicians.

The principal but not the only destructive provision of H.R. 4949 is the requirement that a copy of the examination and correct answers be disclosed after each administration of the test.

Our problem with this provision is that test item specifications to assure relevance, validity and accuracy impose major restraints for a test like the new MCAT.

These specifications were developed by a very extensive effort which began in 1972 and cost the association almost $1.5 million. The initial step in this program was the forging of consensus among medical school deans, admission officers, minority affairs officers, medical school faculty and students, practicing physicians, premedical advisers and faculty of undergraduate colleges about the desired features of the test through national and regional meetings.

Each group presented its own independent views on the objectives to be achieved in a revision of the then existing MCAT test and the characteristics of an examination that would most nearly achieve those goals.

A 3-day conference of representatives from these groups provided a forum to consider various viewpoints and perspectives and develop a master plan for proceeding with the revision.

The master plan served as a guideline for a 2½-year effort, undertaken under contract and supervised by two AAMC appointed advisory committees, one made up of representatives from groups involved in the initial consensus development and the other of experts in the field of testing.

It was not to be an aptitude test which Mr. Weiss seemed concerned about. It was to consist of individual sections devoted to biology, chemistry, and physics, each designed to measure both the level of knowledge achieved in these disciplines by students and their ability to apply this knowledge in problem solving.

Evaluation of problem solving ability in applicants is not only valuable because it is an important attribute of the physician, but it also provides an insight into the native abilities of students beyond their store of memorized facts.

The content of the questions was to be limited to those areas considered particularly relevant to medical education and medical practice and to material covered in first-year college courses in the subjects.

These are the restrictions which we maintain severely limit the number and rate of development of test questions, not as Dr. Haney said, because of the problems of evaluation or equating.

Appendix I shows the high correlations obtained between the views of various groups of judges on the criteria established for the tests.

[Appendix I referred to follows:]

APPENDIX I

Correlations Among Mean Ratings of Topics in Each Science for Various Groups of Judges

	Success in Medical School Criterion A				Success in Medical Practice Criterion B			
Groups of Judges	Biology	Chemistry	Physics	Total Science	Biology	Chemistry	Physics	Total Science
Males vs. Females	.936	.931	.957	.941	.989	.906	.870	.961
Majority vs. Total Minority	.843	.906	.925	.898	.985	.911	.863	.958
Majority vs. American Indian & Spanish Surnamed	.771	.875	.901	.834	.961	.811	.852	.923
Majority vs. Black	.757	.832	.904	.843	.958	.907	.687	.930
American Indian & Spanish Surnamed vs. Black	.624	.799	.893	.734	.900	.765	.635	.872
Pre-Clinical vs. Clinical	.701	.847	.943	.835	.981	.910	.828	.960
Pre-Clinical vs. Physicians in Practice	.805	.895	.934	.890	.971	.916	.897	.961
Pre-Clinical vs. Students	.674	.756	.871	.797	.948	.780	.844	.894
Pre-Clinical vs. Residents	.799	.907	.879	.811	.902	.778	.770	.871
Clinical vs. Physicians in Practice	.837	.792	.930	.870	.953	.788	.787	.926
Clinical vs. Students	.819	.847	.835	.856	.959	.872	.772	.923
Clinical vs. Residents	.802	.787	.811	.662	.903	.644	.548	.824
Physicians in Practice vs. Students	.754	.648	.860	.794	.952	.674	.847	.871
Physicians in Practice vs. Residents	.815	.897	.826	.726	.937	.793	.639	.870
Students vs. Residents	.764	.648	.814	.627	.930	.486	.557	.791

83

Dr. COOPER. From the specifications developed for the test item, an extensive process is followed in preparing questions:

First, they are written by faculty experts from the fields of biology, chemistry, and physics;

Second, they are reviewed by test editors to assure that they meet specifications for relevance and level of difficulty;

Third, they are evaluated in developmental try-out tests involving small representative groups of students who not only answer questions but write their impressions of each question, paying special attention to potential ambiguities and biases and, after a group discussion, they present their findings to our test editors.

They are then submitted to a test committee for review of relevance, accuracy and level of difficulty. They undergo a field trial as a part of a regularly administered new MCAT examination where the answers are not used in arriving at the reported scores of the candidates but to further evaluate level of difficulty of the questions and determine by statistical analysis any previously undetected problems, including ambiguity or bias.

It is only after this very extensive program which occupies 18 to 24 months to develop and validate questions that they become a scored element in the new MCAT test.

We have also developed at each test center a procedure under which students taking the test are able to bring to the attention of proctors any concerns they have about particular test questions.

The proctor has immediate access to the national coordinating center for the test. Students are also encouraged to write about their concerns and criticisms of the test and their comments are carefully considered.

I apologize for this detailed description of how the new MCAT was developed and implemented. I thought it important for you and your colleagues on the committee as you seek "Truth in Legislation" and truth about the possible differences between the new MCAT and more general standardized tests.

I cannot speak for all standardized tests, but the proposed legislation is fatally inconsistent with the establishment and administration of this test.

There are many safeguards to assure the accuracy of the grading procedure and we have great confidence in the system that we have put in place. However, any student concerned about the grading procedure may request a separate hand grading of their answer sheet. If questions persist, they are offered an opportunity to personally observe a third grading procedure.

From its very inception, the new MCAT was designed to provide maximum information to students well in advance of the test. The new MCAT manual, available for $3.25, provides students precise specifications for the subject matter to be covered in three areas tested, biology, chemistry, and physics, the level of difficulty, first-year college courses, and the description of the types of test items used and the rationale of their selection.

The manual walks students through approximately 80 sample questions. Finally, an illustrative test of over 200 questions permits students to evaluate their level of preparation without the need to enroll in a proprietary course.

One of the concerns identified in H.R. 4949 is the appropriate use and limitations of standardized tests. The only way we can see that the bill addresses this issue is to destroy the enormous usefulness of a highly special test, such as the new MCAT test, through its disclosure provision and cause them to be abandoned by medical school admissions committees.

We have an extensive, systemmatic program to inform and educate those using the test results on the limitations and appropriate application of the new MCAT. Detailed technical and practical information is set out in the new MCAT interpretive manual provided to all medical school admissions officers.

In addition, annual, regional, and national workshops are held with admission officers and admission committee members to inform them about the characteristics and proper use of the examination. These workshops also provide for exchange of experiences and information of those using and preparing the test.

To the extent that there are concerns that the new MCAT is the only criterion for admissions decisions, let me reassure the committee.

In a separate study at the University of Washington on admissions process by Hamburgh, and others—Journal of Medical Education, vol. 46, No. 11—the reference is given that new MCAT test results, premedical grades, interview ratings and letters of recommendation from undergraduate faculty members are given approximately equal weight in medical school admissions decisions.

Examination of MCAT scores of applicants accepted to medical schools is consistent with these findings. Between 25 and 30 percent of applicants scoring in the four highest levels of the various MCAT subtests; that is, those with scores from 12 to 15 on the 15-point scale which has a mean of 8, fail to gain entrance to any medical school.

On the other hand, between 9 and 13 percent of applicants with scores in the four lowest levels—that is, scores from 1 to 4—are admitted.

If the quality of the new MCAT is so compromised by Government intrusion that it is no longer of value, what are the alternatives for choosing among applicants?

Should admissions decisions be based on grades earned in college, grades that vary with the individual standards of the professor, the difficulty in content of courses at the same levels in different institutions, on the caliber of other students in the class, on the system of evaluation used by the institution, on the course load the student is carrying, on the extent of his outside work, and so on?

If for these reasons undergraduate grades are considered imperfect criteria, should admissions decisions depend only on subjective impressions of creativity, judgment, determination, capacity for growth or idealism?

Should the choices rest with the letters of recommendation arranged by the applicant?

It does not require long experience on an admissions committee to conclude that the only defensible strategy to fulfill one's responsibilities to the medical school, to all applicants, and to the welfare of the public is to use a mosaic of all of the available criteria of which standardized testing is an important one, in the selection of students.

Since the new MCAT taps so wide a range of knowledge and attributes important in the study of medicine, it serves as a valuable diagnostic indicator of the institutional resources that would be needed to support a candidate to become a well-trained physician should that candidate be admitted.

This is especially important in planning programs for those with educationally disadvantaged backgrounds, and we are proud of the success of the medical schools in maintaining those students in school and graduating them.

It is claimed in the introductory remarks to H.R. 4949 that proprietary preparatory courses use questions from previously administered tests as a part of their review program. No evidence to verify such rumors has been obtained by enrolling knowledgeable MCAT personnel in these courses.

I would also like to note here that our information from questionnaires to students indicate that almost the same percentage of minority candidates take proprietary MCAT preparation course as white candidates.

A major issue is also made about the relationship between parental income and standardized performance. If test performance can be expected to relate to educational opportunity and if educational opportunity is related to parental income, the relationship is not surprising.

The new MCAT test does not correct inequalities in premedical education, but it does furnish an objective basis to call for educational reforms that will equalize this opportunity.

For example, it is not possible to construct remedial programs unless a reliable indication of baseline knowledge is established. Those who purposely or inadvertently promote elimination of standardized testing are burying their heads in the sand, putting the blame in the wrong place, and aiding and abetting the establishment of false standards and the continuation of inadequate preparation for educationally deprived young people.

H.R. 4949 also charges that bias is inherent in standardized tests. For the new MCAT, it is difficult to understand how the universally accepted facts of biology, chemistry and physics, or their application in problem solving, can be biased on any social designation, whether it be racial, sexual, or political.

The Russians tried to inject a political bias into genetics with disastrous results for Soviet genetics under the state-approved Lysenko doctrine. However, we are sensitive to the possibilities of cultural bias and have consistently maintained an extensive review of all test materials by minorities and women.

An analysis of pretest questions for bias has been an important part of question development. Furthermore, performance characteristics of different groups of examinees has been regularly provided to admissions officers as a part of the program of information designed to enhance proper use of the test.

The last sections of H.R. 4949 specify various financial disclosure requirements. We believe that the schedule of fees for the new MCAT that now requires a full day of testing is reasonable and justified on the basis of the services provided.

Our fee structure is regularly reviewed by our executive council, a body with representatives of the deans, faculties, hospital directors, and medical students. They concur in the establishment of all fees.

Though we oppose the passage of legislation regulating standardized testing unless there is a much clearer demonstration of need to extend Government regulation into this area, we find H.R. 3564 less objectionable.

The prohibition of developing test scores based on the performance of other test subjects causes us very significant problems. Unfortunately, this denies admission committees flexibility in relating their criteria to the pool of applicants under consideration.

As we interpret the tests, our views are slightly changed by the testimony this morning and colloquy between Mr. Gibbons and Mr. Goodling. There is also a judicial perception that applicants with highest qualifications on the full range of criteria should receive strongest consideration in medical school admissions decisions.

In view of our careful assessment of the legislation you have under consideration, we strongly urge that this committee not report favorably on either H.R. 4949 or H.R. 3564.

Alternatively, the committee should provide an exception for highly specialized tests such as the new MCAT. We believe that the Congress will be greatly assisted in considering this complex issue by the findings and recommendations of a major study currently underway by the National Academy of Sciences-National Research Council's Committee on Ability Testing.

We are certain that the report of this prestigious and impartial panel will provide valuable insights into standardized testing.

Thank you very much.

Mr. WEISS. Thank you very much.

Our next witness is Mr. Robert J. Solomon, executive vice president, Educational Testing Service.

STATEMENT OF ROBERT J. SOLOMON, EXECUTIVE VICE PRESIDENT, ACTING PRESIDENT, EDUCATIONAL TESTING SERVICE

Mr. SOLOMON. Mr. Chairman, my name is Robert J. Solomon. I serve as executive vice president of Educational Testing Service, where I have worked in various positions for over 25 years.

I came to ETS from a career as a school and college teacher of history and social studies in New York City.

My first job at ETS was in test development, and I helped develop tests used in college and postgraduate admissions.

I am grateful for this opportunity to talk to you today about further proposed changes in testing.

We have submitted written testimony with attachments. I welcome this opportunity to make some additional remarks.

I should first like to summarize the position of ETS regarding the legislation that is being proposed.

The proposed legislation reflects the legitimate concerns of students, parents, educators, and all those interested in educational policies and practices. We are sympathetic to those concerns and understand the need to respond to them.

We believe that in the past few years we have responded in some measure to the concerns. The discussion and debate stimulated by proposed legislation in some States and in the Congress indicates, however, that more needs to be done. We are prepared to do more.

We do not believe, however, that there is a convincing case for Federal legislation at this juncture. The introduction of new elements of control over education should not be undertaken without utmost caution.

Some provisions of the proposed legislation will create serious problems that will affect negatively the quality, cost, and services of testing programs. We recommend several actions that the Congress, the higher education community, and the testing agencies should first consider.

We hope, and we are prepared to initiate constructive efforts to address and resolve the issues with which we are all concerned, including disclosure of the contents of major admissions tests insofar as we are able within the interrelated constraints of quality, service, and cost, we hope time and events, including our actions, demonstrate that Federal legislation is unnecessary.

Legislation recently enacted in the States of California and New York, and legislation newly proposed by Congressman Gibbons of Florida, H.R. 3564, and Congressman Weiss of New York, H.R. 4949, have stirred a good deal of public discussion about the responsibilities to students of testing organizations such as ETS and the colleges they serve.

Many things have been said in the heat of debate by critics of testing that are untrue or misleading, arising no doubt from misunderstanding about a complex subject too little understood by the public at large.

The main concern of ETS in this debate is to assure that those who may make legislative decisions about testing understand the probable impact of those decisions on students, on educational institutions, and especially on the capability of testing organizations to maintain the quality of the tests themselves.

Educational tests, especially today when grade inflation and lack of consistency in grading standards from school to school are widespread, constitute very important instruments for maintaining the quality of educational programs. In an effort to increase the information about tests available to students, a goal we at ETS share, we must also be careful not to damage the quality of tests and related educational standards.

It is our hope that open discussion of many aspects of testing will dispel some of the misunderstandings about testing and that out of such discussions as these new initiatives can be planned in an atmosphere of cooperation and problem solving which will benefit all parties concerned, students colleges and developers of tests.

One source of misunderstanding is the relationship of ETS to the testing programs we administer. These relationships are important in order to understand the context within which this testimony is offered. Virtually all of the admissions testing and related services conducted by ETS are done so under agreement with organizations such as the College Board and the Law School Admission Council. [See attachment 1 for summaries of College Board and Law School Admission contract.]

Each of these policy groups is representative of the educational constituency it serves. These membership organizations, their processes of election and their governance are independent of ETS.

The decision to use ETS services rests, in each instance, with the educational institutions and their representatives. Moreover, the ultimate judgment of the quality and utility of testing programs rests with the educational community that uses the services.

On a day-to-day basis, ETS is operationally responsible for the development of tests for these client groups, and for the administration of tests, scoring the answers, and reporting the scores. In addition, ETS undertakes research and development necessary for the creation of valid and reliable tests, and the writing and production of most but not all program publications.

Although ETS staff may provide information and advice on matters under consideration by these policy groups, final decisions are made by them. For these reasons, I can speak for ETS, but I cannot speak for the independent bodies we serve, whose policies are determined by them.

ETS is not only accountable to its client groups, but is also ultimately responsible to a board of trustees drawn largely from the leadership of the educational community. [See attachment 2.] We have provided a list of those trustees for the last 5 years. This body provides oversight for the activities of the organization and is a critical source of external direction and advice.

ETS was established by the educational community about 30 years ago as a nonprofit organization to serve education in the field of testing and research. However, its roots go back to the turn of the century when the College Board was first formed in 1900. The situation that then confronted students was chaotic, each college had its own entrance examination, often covering different subject matters.

The development of a uniform system of college entrance examinations did much to bring order out of chaos and extend freedom of choice for students.

The ensuing emergence of standardized tests designed to assess fundamental outcomes of education, such as reading comprehension, verbal reasons, and mathematical ability, such as the College Board's scholastic aptitude test, helped to advance two broad social goals, both of which were realized to a remarkable degree beginning in the 1930's and 1940's:

First the enrollment of students from secondary schools unknown to the colleges, enabling colleges to broaden their student bodies eventually to include an economic and social mix unprecedented in American higher education, and;

Second the development of the comprehensive high school, free from the narrow tyranny of the traditional college preparatory curriculum, and organized to offer a wide range of programs to students of diverse backgrounds and interests.

In the 1940's and 1950's, the national testing programs developed for various membership associations were designed primarily as a means of helping colleges, universities, and professional schools make the best possible match between the institution and the students who sought admission. The design of programs at that time reflected the prevailing view that educational decisionmaking was primarily an institutional prerogative in which students themselves had relatively little voice.

The 1960's and 1970's have seen a reevalution of traditional institutional attitudes regarding the role of students and, additionally, major changes in admissions testing and related services. It was only 25 years ago, for example, that the membership of the College Board voted to provide scores on the SAT to students.

In the past decade particularly, there has been a remarkable increase in the quality and quantity of information available to students about admissions tests.

For example, full copies of recent forms of each of the major admissions tests that ETS develops are routinely made available by sponsors of testing programs to students, and we have submitted copies of the College Board, Law School Admission Council, Graduate Management Admissions Council, and the Graduate Record Examination Board. [Attachment 3 has been placed in subcommittee files.]

Indeed, our goals and those of the sponsors of the tests have been threefold:

One, to increase the flexibility and ease with which students and institutions can obtain needed testing services and related information;

Two, to provide more and better services tailored to the diverse needs of special groups of students.

For example, tests in Braille and large type for visually handicapped students, special administrations of tests for students whose religious beliefs preclude Saturday attendance, and special administrations of tests on military bases for servicemen.

Also, you will be pleased to know over the past 25 years the fees for SAT, GRE, LSAT and GMAT have increased typically only about 47 percent, as contrasted to a 191 percent increase in the Consumer Price Index for the same period. [See attachment 4.]

The record of ETS attests to its commitment to the goals which underlie the intent of laws recently enacted in California and in New York, and the bills introduced into the House by Congressmen Weiss and Gibbons.

We do provide a broad array of information to students about tests and their appropriate uses. Dozens of publications are made available to students, schools, and colleges which contain most of the information called for in these bills.

We take scrupulous care to see that the construction and administration of tests are fair, reliable, valid, and unbiased.

We respect the rights to privacy of individuals, and have developed very detailed procedures for insuring the confidentiality of data retained in our operational and research files.

Nevertheless, we have in the past opposed bills such as the one recently enacted in New York because many of their special provisions could inadvertently compromise these goals or otherwise undermine the quality of the testing programs.

However, I do not wish to be misunderstood. ETS, working together with the test program sponsors, will make every effort to comply with the provisions of the legislation. At the same time, we will look toward clarification of the law and the possibility of constructive amendments that may yet correct some of the unfortunate and, we believe, unintended consequences of this legislation.

We have listened carefully to the debate concerning the disclosure of test questions and answers to students, and we have engaged in earnest conversations on this issue with many people and groups throughout the Nation over the past year. We are sensitive to the arguments for extending the scope of our services to include disclosure of test questions and feedback on individual student answers. There are serious problems that must be solved in meeting this new demand for service. Nevertheless, we are committed to finding ways to satisfy the understandable desire for such information, while at the same time maintaining the quality of our tests. To do this we need time, patience, and a climate of cooperative, open exchange—not the accusatory and dogmatic climate that has clouded the debate up to this point.

We have submitted specific comments on H.R. 3564 and H.R. 4949. The provisions of the latter bill would make it extremely difficult, if not impossible, for us to continue to provide current levels of quality, price, and service to all students. [See attachment 5.]

These difficulties are illustrated in the case of the Graduate Record Examinations and the Graduate Management Admission Test in an additional submission that we have made available to this committee. [See attachment 6.]

It is important to keep in mind that the reason test development costs have been a small proportion of total program costs is because of the ability of program sponsors to reuse test forms. That would not be so under H.R. 4949.

For example, in the graduate record examinations program, in addition to 5 regular Saturday administrations and [illegible] Sunday administrations, there were this year 116 administrations in 9 large urban areas.

We don't see how such administrations could continue under the provisions of H.R. 4949. Also, H.R. 4949 will make impossible the score equating methods now used for all the major admissions tests administered by ETS except the Scholastic Aptitude Test.

To force a change in this essential element of national testing programs represents a major disruption that affects all aspects of the examinations, including not only the score equating but, also, the content and structure of the examinations.

Such a change demands careful study and research. In addition, although it is difficult to demonstrate before the fact, the obligation to multiply the development of new tests severalfold in a relatively short time would threaten the quality of the tests, which is essential to their ultimate validity.

The development of good standardized tests is a painstaking, time-consuming process that requires the combined talents of technical specialists and scholars.

It would be unfortunate if, in the haste to rectify apparent shortcomings of testing, we believed that legislating disclosure of test questions was a panacea.

There is no better illustration of the need for clarification, rather than rhetoric, than in the question of whether admission tests are culturally biased. This is an important question, which ETS, test sponsors, colleges, and researchers throughout the country have studied over the years, because there is a shared concern over the substantial variation in average test performance among groups.

For instance, while scores for whites and disadvantaged minority groups overlap, a typical result is to find that only 10 to 20 percent of disadvantaged minority groups score above a point that is average for whites, that is, exceeded by 50 percent of whites.

These differences will not surprise anyone familiar with the inequalities in the social, economic, and occupational spheres of American life. Although many special educational programs have been developed at the Federal, State, and local levels to repair the effects of educational and social neglect, these programs and other measures cannot change so quickly what years of malpractice have perpetuated.

Nevertheless, the argument is made that tests of academic ability are biased because they represent middle class culture. These tests do reflect skills and knowledge considered important in the mainstream culture, in many jobs, and in higher education. But the fact that a test mirrors the common culture is a poor basis for calling the test biased.

Investigators who have studied the issue agree that one commonly accepted way to evaluate whether or not admissions tests are discriminatory is to determine whether a difference in test scores between groups is also reflected in a corresponding difference in later performance in college or graduate school. A number of such studies have been undertaken on our tests, both at ETS and by researchers elsewhere.

The results have been quite consistent. Differences in test score averages across groups are consistent with differences in actual performance in college. Tests typically predict the same way with the same validity for whites as for minorities.

One can also ask whether certain types of test questions are discriminatory. Test makers must avoid questions that give an unfair advantage to one group over another. Test questions undergo formal review by minority staff members or outside reviewers, and are examined for possible cultural bias.

Similar reviews are made with respect to sex bias. After some years of experience and careful attention, we feel we are reasonably successful in keeping biased items out of tests.

In addition, there are research techniques for evaluating whether individual questions are relatively more difficult for one group or the other. Sometimes these differences are not easily explained. For example, when compared with whites, black students tend to score somewhat lower on mathematical than verbal test questions, though the former are ordinarily considered more "culture-free." Also, women are inclined to do somewhat less well than men on geometry questions, but that fact is more a reason for reforming society than for reforming testing.

In his statement introducing H.R. 4949, Congressman Weiss referred to a 1977 study by the California Postsecondary Education Commission reporting a difference in average income for high-scoring and low-scoring students. Our data confirm that there is such a statistical relationship. In fact the particular figures cited evidently came from a report prepared by ETS for the College

Board. Such information has been widely distributed to schools and colleges through annual reports since 1972.

The fact that SAT scores are related to students' family incomes is often misinterpreted or misunderstood. The relationship is far from perfect and shows up only in the averages for groups of students. Many individual students from low-income families attain high scores and many individuals from affluent families attain low scores.

While the test does show differences between students from low- and high-income homes, and those differences are distressing and significant, it is important to recognize this as part of a larger pattern. It is unfortunately true that many measures of growth and development show similar relationships to socioeconomic level.

We know from the extensive research of Project TALENT that socioeconomic measures of family status have a definite relationship to whether students keep up with their age group or fall behind in grade placement in school. Furthermore, research evidence indicates students from poorer homes make somewhat lower grades in school on the average. Our own studies in individual high schools indicate that both test scores and grades show a similar low positive relationship to family income.

Income is also related to other vital signs. In families of low socioeconomic level, infant mortality is much higher. Life expectancy is some 8 to 10 years lower in the lowest as compared to the highest economic class.

Do these statistical relationships between income and other important measures of personal development of quality of life indicate that those measures are biased against poor people? I think not. Such data simply support the well-known fact that we still live in a society where there are shameful discrepancies in health, growth, and achievement resulting from unequal opportunity. Congress is well aware of these social inequities and has sought through vigorous legislative programs to mitigate the effects of economic deprivation.

Tests like the SAT help to bring these problems into focus because they provide a standard yardstick of developed cognitive ability. Tests measuring developed abilities of proven importance to college work are a great strength to a fair selection system because they measure skills relevant to intellectual demands on the student, they insure that all students meet the same standard, and they can be given to large numbers of people.

Nevertheless, most colleges and universities are properly reluctant to rely exclusively on admission tests and grades in selecting students. Admissions officers appreciate that even though these measures are the best predictors of academic performance, the relationship is certainly not exact. Furthermore, institutions have a variety of social and educational objectives for which they may consider other student qualities especially relevant—for example, special talents, experience, accomplishments.

The outlook for additional measures that are relevant to selection decisions is promising. For a long time, ETS has argued against undue reliance on test scores and in favor of a broader definition of talent in American society. We have been engaged in a variety of projects that will enable educators and others to give more attention to the varied skills and accomplishments each individual has to offer. These include:

A cooperative study by College Board, ETS and nine colleges and universities to improve the assessment of personal qualities in admissions.

A study on the development of nontraditional methods of assessing experiential learning. This work resulted in the formation of CAEL—The Council for the Assessment of Experiential Learning in Columbia, Md.

An experiment by the Graduate Record Examinations Board with an Inventory of Documented Accomplishments to accompany a student's application to graduate school.

The answer to real educational and job opportunity lies not in setting aside careful objective assessment of developed abilities. Rather we must be open to all forms of talent, value the different contributions different individuals can make, and take the time to assess those capabilities fairly.

It is our hope that legislation, if it is needed, will grow out of serious study of the facts and not out of misunderstanding and simplistic rhetoric. At stake, we feel, is the entire system of college admissions which has served this country well. With new understanding, and vigorous search for fresh approaches, testing can serve our Nation even better in the future.

As a means of fostering a new beginning, I should like to reference [see Attachment 7] and call your attention to the principles which ETS believes should guide the exercise of its own responsibilities to students. These responsibilities also extend to the larger society that ultimately must judge the value of our work. During the past few years, ETS has developed guidelines [1] for insuring that its work conforms to uniformly high standards in seven areas of basic importance: Accountability; confidentiality of data; product accuracy and timeliness; research and development; tests and measurement; technical quality of tests; test use; and technical assistance, advice, and instruction.

ETS further recognizes that the student has a concern in all aspects of admissions—but especially in the quality and character of service provided by testing agencies and in the valid and fair use of tests and other information by schools and colleges. We believe that these two closely related areas of concern on the part of students can be met more effectively if, to the maximum feasible degree, the products of the testing programs and the admissions processes of schools and colleges are open, fully understood, and stripped of whatever unnecessary vestiges of mystery may yet obscure them from students and the public. As this committee considers H.R. 3564 and H.R. 4949, we respectfully urge that you consider your actions in the light of the effectiveness of testing agencies and test users in these regards.

As a demonstration of our good faith, we will undertake certain initiatives and propose others for consideration:

First, in response to the New York legislation, ETS is engaged in serious and intensive work to determine how we can comply with the law, maintain the quality and integrity of the tests, and hold service reductions and cost increases to a minimum. Fifty members of our professional staff are currently involved in nine task forces studying the problems we face and identifying possible solutions and tradeoffs. We are scrutinizing every aspect of our work, including test development, test administration, equating and other psychometric concerns, and overall operational implications. The results of our work—and those of other organizations similarly covered by the New York law—will prove helpful in light of Federal concern. We recommend that this committee view the New York experience as an opportunity to assess and evaluate the impact of such legislation, rather than rush headlong into national legislation with all of the attendant risks for serious damage that this course of action will incur.

---

[1] *Principles, Policies and Procedural Guidelines Regarding ETS Products and Services*, Educational Testing Service, Princeton, N.J. (Rev. February, 1979).

In addition, we would recommend that the National Academy of Sciences, which is currently conducting a study of ability testing, examine the particular set of concerns which have led to this hearing. The Academy would assess the benefits and costs, both technical and practical, of possible solutions, and report in a fixed but reasonable period of time on the effects of the California and New York legislation and on what testing agencies have done in these two States. This would provide a sorely needed independent review of the need for and impact of legislation. We would cooperate fully with such an undertaking.

In addition we will urge that the American Council on Education, or an organization similarly representative of higher education, appoint a Commission on Test Use to establish guidelines for the proper use of tests by colleges and universities. The ACE has taken similar leadership to strengthen self-regulation in other areas of higher education. This approach is consistent with higher education's willingness to regulate itself as a means of resolving educational problems.

Also, we will sponsor together with all other interested parties an open working conference on testing legislation. The purpose of the conference will be to examine the potential problems of overlapping legislation, the need for model legislation, and the question of preemptive Federal legislation. Such an undertaking will require the participation of students, legislators, lawyers, and educators.

Finally, to further underscore our commitment to the principle of providing greater information to test takers, ETS will recommend at the fall meeting of the National Teacher Examinations Council—the body which establishes the policies for operation of the NTE—that by 1980-81, or as soon as possible thereafter, the contents of the NTE common examinations be released to test takers for at least one of the three national administrations each year. The NTE dates back to the late 1930's and is used in teacher certification and selection. H.R. 4949 does not affect NTE, since the tests are not used for postsecondary admission. Nevertheless, we believe that beginning teachers would benefit from having more information about their performance on the examination. Further, we would propose that we work cooperatively with the two national teacher organizations, the American Federation of Teachers and the National Education Association, and also the American Association of Colleges for Teacher Education, to create materials to accompany the disclosed test content which would assist beginning teachers to understand better their performance on the examination.

One final observation. Admissions tests for higher education, such as those provided by ETS, require a mutual appreciation on the part of all concerned parties of the special relationships among the public's right to know, the university's need for educational autonomy, and the testing organization's obligation to sustain high professional standards in its work. We believe that policies consistent with these principles will further insure that testing services continue to serve the public interest. If, in the process, we can increase public understanding of admissions testing by students and parents, our entire educational system will benefit.

Mr. WEISS. Thank you very much. We appreciate the testimony which all of you have given. Apparently because there are a number of suspensions on, we will be over there for at least 20 minutes, perhaps as much as a half hour. What I would suggest that you do is to allow you to catch your breath and we will resume when the vote is over, which will be in about a half hour.

Thank you very much.

[Recess.]

[Attachments to Robert Solomon's statement follows:]

ATTACHMENT 1

The attached document summarizes the agreement dated July 1, 1974 between the Law School Admission Council (LSAC) and Educational Testing Service (ETS). This summary is provided on request to those who wish to know about the contractual framework within which these two organizations operate. Individuals who would like to check the summary's accuracy may inspect a copy of the agreement (without specifications schedules) at one of the following locations by contacting in advance the indicated official:

John S. Kramer, Esq.
General Counsel
Educational Testing Service
Rosedale Road
Princeton, New Jersey 08540

Dean Frederick M. Hart
President, Law School Admission Council
School of Law
University of New Mexico
1117 Stanford Drive, N.E.
Albuquerque, New Mexico 87106

AGREEMENT DATED JULY 1, 1974 BETWEEN THE LAW SCHOOL ADMISSION COUNCIL AND EDUCATIONAL TESTING SERVICE

SUMMARY

A. DEFINITIONS

Certain specific terms used in the agreement are listed and defined in the first section.

B. ETS SERVICES AND ACTIVITIES

ETS's responsibilities include all services and activities reasonably necessary, incidental or customary to the operation of the LSAC program, including test development, research, test administrations, LSDAS (Law School Data Assembly Service), validity study services, candidate referral services, auxiliary publications, and collection of candidate fees and income.

All services and activities are to be performed to the reasonable satisfaction of LSAC.

It is recognized that LSAC and ETS may enter into separate agreements with one another for additional services. Unless otherwise specified, such separate agreements automatically incorporate applicable provisions of the agreement with the exception of compensation and payment provisions.

C. POLICY DIRECTION

LSAC determines the general policy governing the conduct of the services and activities, including the setting of fees and other charges made to candidates.

D. COMPENSATION AND PAYMENT

ETS receives for its services a "basic price" for each LSAC program year, which consists of the aggregate of actual contract costs plus applicable fees, or a "ceiling price", whichever is lower. The ceiling price for each program year is the aggregate of the bid prices for that year, determined as follows:

1974-75	Based on 1973-74 actual costs, adjusted for projected volume estimates, plus fee;
1975-76	Bid prices for all services (detailed in a Schedule attached to the agreement);
1976-77	Based on 1975-76 actual costs, adjusted for projected volume estimates and inflation, plus fee; or the basic price paid to ETS for 1975-76 (adjusted as above), whichever is lower;
1977-78	Based on 1976-77 actual costs, adjusted for projected volume estimates and inflation, plus fee; or the basic price paid to ETS for 1976-77 (adjusted as above), whichever is lower.

The agreement also provides for an incentive fee to ETS in any program year that actual costs are lower than the ceiling price.

Basic research and development costs, on which ETS receives no fee, are included in the above contract costs. However, other research and special projects must be separately agreed to by LSAC and ETS.

After the fourth operational year (1977-78), the ceiling price is to be determined by written agreement between ETS and LSAC. In the absence of such agreement, unresolved matters will be settled by arbitration, as provided in a separate section of the agreement covering arrangements for such arbitration.

E. FINANCIAL ACCOUNTS

All monies collected from candidate fees and other sources of LSAC program income are the exclusive property of LSAC and are deposited in a separate LSAC program bank account, over which LSAC has sole control.

F. REPORTS AND AUDITS

ETS submits to LSAC monthly financial statements covering the LSAC program, and also provides appropriate performance data. LSAC may conduct an audit of ETS books related to the LSAC program and, in addition, receives a certified copy of an annual independent audit of ETS for each year the agreement is in effect.

G. TERM OF AGREEMENT

The initial term of the agreement is from July 1, 1974 until June 30, 1978, subject to automatic renewal after that date, unless terminated as provided under termination provisions.

H. OWNERSHIP, COPYRIGHT AND OTHER RIGHTS

LSAC owns, controls, and, where applicable, has copyright, tradename, or patent rights in: titles and acronyms used now or in the future in the LSAC program; publications developed primarily for the LSAC program; reports or other work product of research produced primarily for the LSAC program; materials, forms, data, computer systems and software developed primarily for the LSAC program; and test items, test forms and answer sheets first developed or used primarily for the LSAC program. In addition, LSAC may request a royalty-free license to use certain ETS-owned retired Law School Program test materials, and ETS may request of LSAC a similar license to use retired LSAC test materials developed by ETS. ETS may also use for other purposes, computer systems and associated software developed primarily for LSAC, upon notification to LSAC of its intent to do so.

I. RESEARCH

ETS insures that principal researchers named in proposals are allocated sufficient amounts of work time to effect timely completion of such projects, and that no such principal researchers are reassigned prior to such completion without written authorization of LSAC.

Written reports are submitted to LSAC on the progress of and costs incurred in connection with research conducted by ETS under this or separate agreements with LSAC.

J. TERMINATION

On or after June 30, 1978, the agreement may be terminated by either party upon written notice to the other party by June 30 of the preceding year. After notice of termination, ETS is to make reasonable efforts to reduce or eliminate any services or activities connected with the program when LSAC so requests in writing. Provision is made for an orderly transfer of responsibilities to a new contractor, if such a decision is made.

K. CONSULTATION AND AMENDMENT

ETS and LSAC will consult freely with one another in order to resolve any problems or difficulties which may arise in connection with the LSAC program, and prior to initiating any arbitration or litigation relating to the agreement. Either party may propose an amendment to the agreement. In addition, if, in the judgment of LSAC, it is desirable to reduce, add to, or change any of the technical or operating aspects of the LSAC program, or to cancel any services or activities other than LSAT administrations and LSDAS (which may be cancelled only by amendment to the agreement), and the parties are unable to agree on an amendment covering such changes or cancellations, then LSAC may direct such change or cancellation on 60 days written notice to ETS, and ETS will diligently comply with any such direction. Unless otherwise agreed, the effect upon price for services will be settled by arbitration. If, in the judgment of ETS, the direction involves a question affecting its professional or ethical responsibility, then ETS also has the right to submit to arbitration the question of whether it has such effect and whether ETS must comply with the direction.

L. SECURITY AND CONFIDENTIALITY

ETS is required to implement security and test administration procedures to insure security of test content, proper identification of candidates, and related matters. Any breach of security is investigated and reported to LSAC.

All data collected, generated, or stored by ETS primarily for the LSAC program is to be stored by ETS separately from data collected for other purposes, so as to insure complete confidentiality, retrievability, and destructibility at all times. While ETS has custody of such data, the handling and storage will be in accordance with the ETS policy on confidentiality of data, and ETS security procedures. LSAC likewise agrees to maintain the confidentiality of such data.

M. SUBCONTRACTS AND OUTSIDE RESEARCH

ETS may subcontract work in connection with the LSAC program, but remains responsible at all times for proper performance of the work. LSAC retains the right to arrange for performance of research projects, special studies, and publications by third parties.

N. ARBITRATION

Any claim or controversy arising out of or relating to the agreement will be settled according to the rules of the American Arbitration Association, or such other rules as the Association may specify.

O. IMPOSSIBILITY

A limitation of liability is provided for each party in the event it is unable to perform its responsibilities due to circumstances beyond its control.

P. NOTICES

Designation of the recipient and address is provided by each party for any notices concerning the LSAC program or the agreement.

Q. ASSIGNMENT

Neither the agreement nor any interest in it may be assigned or transferred by LSAC or ETS without the written consent of the other party.

R. PRIOR AGREEMENTS

With the exception of certain agreements covering research projects, all prior agreements between LSAC and ETS are terminated.

S. APPLICABLE LAW

The agreement is to be construed and enforced in accordance with the laws of the State of New York.

SCHEDULE 1

Contract specifications for the program year 1974-75 are listed, including detailed descriptions of the services to be provided, prices for services, budgets, and similar information.

SCHEDULE 2

Contract specifications as described under Schedule 1 above are listed for 1975-76 and beyond.

ATTACHMENT 2

MEMBERS OF ETS BOARD OF TRUSTEES 1974 - 1979

(Affiliation as of the Time of Trusteeship)

Elie Abel
Harry and Norman Chandler Professor
Department of Communication
Stanford University

Robert E. Bell
Principal and Vice Chancellor
McGill University

M. Anne Campbell
Commissioner of Education
Nebraska Department of Education

Thelma T. Daley
Career Education Specialist
Board of Education of Baltimore County

John B. Davis, Jr.
President
Macalester College

Alfredo G. de los Santos, Jr.
Vice Chancellor for Educational Development
Maricopa County Community College District

Joseph W. Fordyce
President
Junior College District of St. Louis

F. Sheldon Hackney
President
Tulane University of Louisiana

John W. Hennessey, Jr.
Professor of Management
The Amos Tuck School of Business Administration
Dartmouth College

Kurt M. Hertzfeld
Treasurer
Amherst College

Roger W. Heyns
President
American Council on Education

Wayne H. Holtzman
President, Hogg Foundation for Mental Health
Professor of Psychology and Education
The University of Texas at Austin

Clark Kerr
Chairman and Staff Director
Carnegie Council on Policy Studies in Higher Education

Robert J. Kingston
President
The College Board

Juanita M. Kreps
Vice President
Duke University

C. Peter Magrath
President
University of Minnesota

Sidney P. Marland, Jr.
President
The College Board

Peter Masiko, Jr.
President
Miami-Dade Community College

Gertrude G. Michelson
Senior Vice President
Macy's, New York

Charles D. O'Connell, Jr.
Vice President and Dean of Students
The University of Chicago

Jack W. Peltason
President
American Council on Education

John A. Perkins
Vice President--Administration
University of California at Berkeley

MEMBERS OF ETS BOARD OF TRUSTEES 1974 - 1979, continued

A. Kenneth Pye
Chancellor
Duke University

Sister Joel Read
President
Alverno College

Wilson C. Riles
Superintendent of Public Instruction
State of California

Paul F. Sharp
President
The University of Oklahoma

Virginia B. Smith
President
Vassar College

Stephen H. Spurr
Professor, Lyndon B. Johnson School
of Public Affairs
The University of Texas at Austin

William W. Turnbull
President
Educational Testing Service

Albert N. Whiting
Chancellor
North Carolina Central University

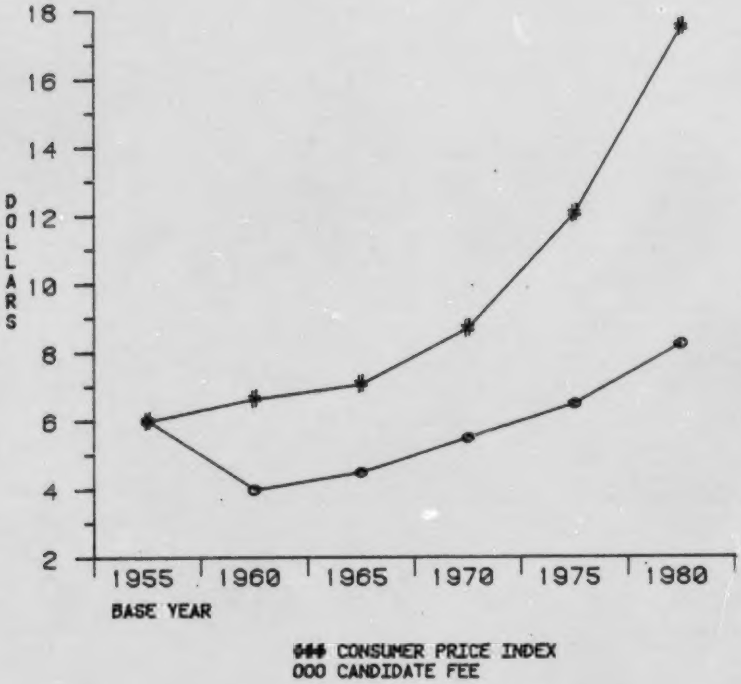
Charles E. Young
Chancellor
University of California at Los Angeles

ATTACHMENT 4

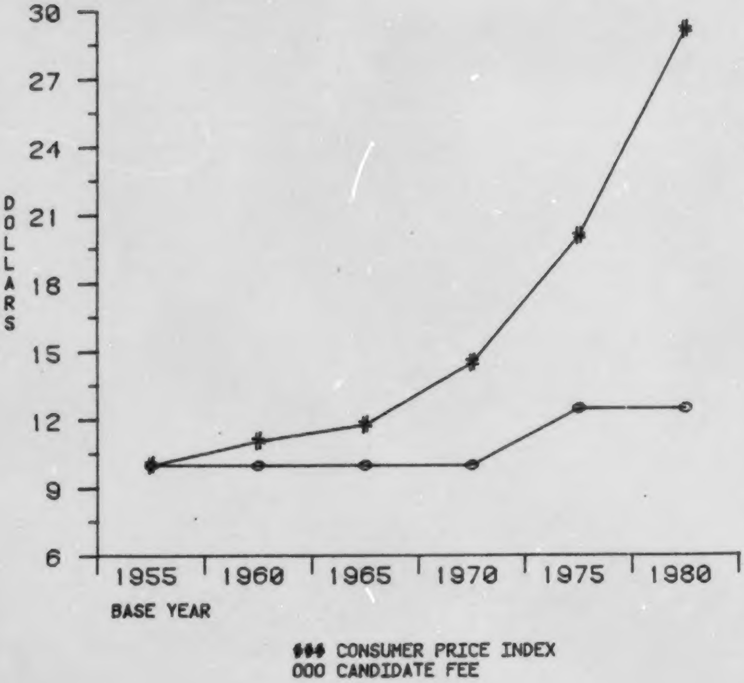
HISTORICAL CHANGE OF FEES VS. INFLATION

	1954-55 Fee	Inflation @ 191.68%	1979-80 Actual Fee
Admissions Testing Program	$ 6.00	$17.50	$ 8.25
Graduate Management Admission Test	10.00	29.15	12.50
Graduate Record Examination	8.00	23.33	14.00
Law School Admission Test	10.00	29.15	15.00

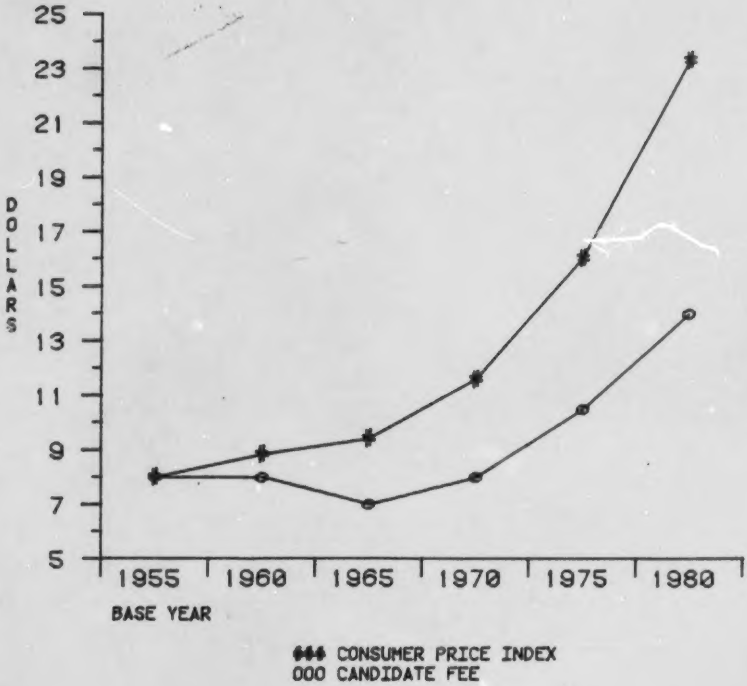
ADMISSIONS TESTING PROGRAM



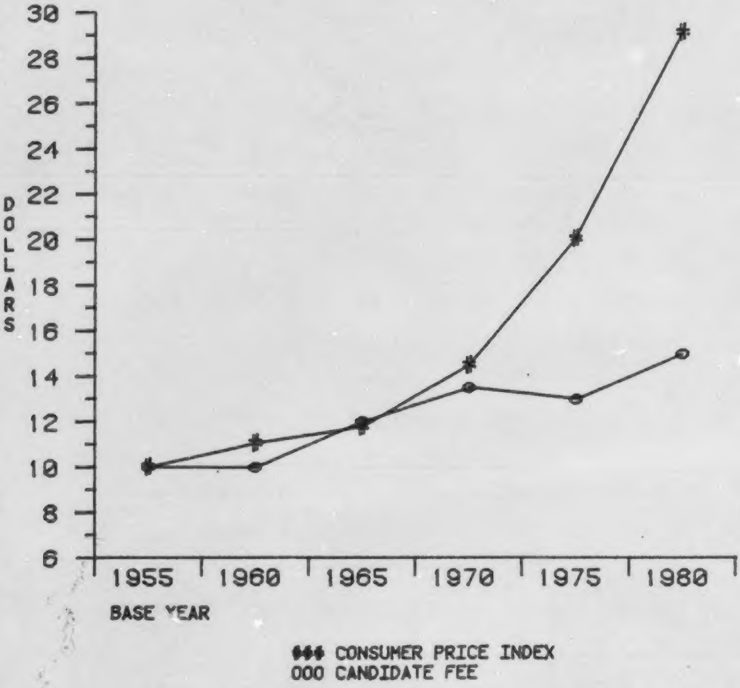
GRADUATE MANAGEMENT ADMISSION TEST



GRADUATE RECORD EXAMINATION



LAW SCHOOL ADMISSION TEST



EDUCATIONAL TESTING SERVICE
PRINCETON, NEW JERSEY 08541
COMMENTS ON H.R. 4949 AND H.R. 3564
JULY 31, 1979

INTRODUCTION

ETS has reviewed the two legislative proposals and has identified the following major areas of concern which appear to us to create problems which could be avoided.

H.R. 3564

Most of the significant problems in this legislation are related to Section 6, and those problems are of four kinds:

(1) There can be problems of interpretations because of the requirement that (for example) "each applicant...shall be notified" (Section 6(a)(1)) or "each individual...shall be notified." (Section 6(b)) without specifying a locus of responsiblity for providing or notifying.

(2) Under Section 6(b)(2), an individual must be informed as to how she/he "ranked on total test performance" as well as "specific performance in each of the subject or aptitude areas tested." This requirement for "total test performance" would conceivably force the reporting of a total score for such tests as the GRE Aptitude Test (in addition to the currently separate Verbal, Quantitative, and Analytical scores) despite professional judgement that such a Total score would be meaningless or subject to gross misinterpretation.

(3) The requirement in Section 6(b)(3) that individuals be informed of "the score required to pass the test for admission to such occupation or the score which is generally required for admission to institutions of higher

education" creates two problems. First, for most admissions activities, ETS and other testing agencies already state publically that there should not be a score "required for admission." ETS would have no way of knowing what any such score might be were it (wrongly) required by an institution. Second, even if ETS and other testing agencies knew of each specific "required" score - whether for an institution of higher education or for admission to practice in an occupation in a given state - it would still be impossible to report that "score which is generally required" as opposed to each specific requirement.

(4) Finally, the requirement (Section 6(c)) that no test "shall be graded... on the basis of the relative distribution of scores of other test subjects" is unclear as to whether the current scoring system for all national tests - in which an individual's score is presented as a number on a score scale developed through reference to an original population which took the test, and is frequently interpreted though "percentile ranks" which compare that performance with the performance of other individuals - would be prohibited.

H.R. 4949

Generally speaking, this bill presents problems of definition that go to the very heart of testing, and a major effort to clarify terms early in the legislative process will avoid misunderstandings at a later time, should the bill become law.

Another general problem area of the bill is the locus of responsibility for the different sections and subsections.

It needs to be determined whether in the filing requirement and the public record aspect of this bill there is a possibility that it represents a position that is in part in direct conflict with provisions of the Copyright Law of 1976.

More specific concerns include:

(1) With one exception, the result of ambiguous language, ETS supports the Findings and Purposes. Section 2(a)(3) requires that "standardized tests should equally reflect the academic ability of students, regardless of their ethnic, regional, racial, or economic background." If that statement means that tests should not be systematically biased against any student or group of students because of ethnicity, race, or economic background, then ETS supports the statement. If, on the other hand, it means that tests should necessarily produce equal mean scores for members of such identified groups, or equal mean scores for all test takers, then the statement represents an unreasonable and untenable position.

(2) The current language in Section 3 makes an erroneous assumption about the amount and kinds of information that ETS or other testing agencies gather about individuals and institutions. For example, a testing agency will not know, and therefore cannot report, the entrance criteria that institutions use or what factors are weighted in determining admissions at those institutions, the income levels of all people who take the tests, or the careers which students in varying academic programs will enter. Furthermore, while ETS completely supports the concept of prompt score reporting, the language in this section fails to recognize that there

are numerous reasons -- including student error, mail delays, or investigations of possible cheating -- which would routinely prevent timely reporting of all scores.

(3) Section 4 requires ETS and other testing organizations to obtain and provide the Commissioner with copies of any studies based upon data we provide. Because ETS "provides data" each time we report scores to institutions or departments within institutions, we could be held responsible for obtaining and reporting to the Commissioner any study based on those data, even though we might not be aware of the study's existence. In a related problem, even when we consciously provide data for a research study (such as those conducted by the National Science Foundation), it would be difficult for ETS to require that the other organization make its report public. Also, the legislation as now written does not distinguish between preliminary (draft) and final reports prepared by ETS, or distinguish between reports which can reasonably be made public or proprietary data that may be summarized. In addition, while this section protects individually identifiable data, the confidentiality of institutional data is not adequately protected. This section also requires a disclosure of contracts, despite the fact that ETS and other testing organizations frequently compete for contracts to compare and administer tests. This section thus creates problems concerning appropriately proprietary information for all testing agencies. Under the Freedom of Information Act, of course, any pertinent government contracts are now available.

(4) In Section 5, copies of individual tests, and of students' answer sheets and the correct answers, must be provided. While ETS can support the general concept of test disclosure for the purposes outlined in the legislation, the current bill as drafted does not provide for several exclusions which appear to be desirable for the very test-takers that the bill is designed to aid.

One of the most difficult problems posed by the proposed legislation is in the area of score equating. Admissions tests are offered many times during the year, and fairness demands that a score earned in October be comparable to a score earned in February, and a score earned in 1977 must be comparable to a score earned in 1980. Strict interchangeability of scores can be achieved only by precise score equating. Three different methods of equating are employed at ETS in the testing programs affected by the proposed legislation.

The first, used on the Scholastic Aptitude Test, involves giving the same equating questions to two groups as part of the two test editions to be equated. A second method, generally used in the GRE Aptitude Test, the Graduate Management Admission Test, and the Law School Admission Test, involves administering two or more test forms (separate editions of a test that meet the same test specifications) administered in alternating order. A third method, generally used in the College Board Achievement Tests and the GRE Advanced Tests, involves imbedding questions from previous editions of the same test in a current edition.

All of the methods of equating rely on the use of previously used but secure (not disclosed) test questions whose statistical characteristics, such as the level of difficulty, have been firmly established. The equating method used in the SAT can still be employed under the proposed legislation, since sets of equating questions that are not used in calculating the student's scores do not have to be disclosed.

Nevertheless, even for the SAT, the disclosure provisions will create substantial problems for all tests used in special administrations where the students are neither sufficiently numerous or representative to ensure reliable equating.

For the other major tests, however, it will be impossible to use the equating methods that are in use now and which have been employed in these programs for decades. To introduce a different equating method, such as the one used for the SAT, represents a major change that affects all aspects of the program, including publications, test administration and timing, and test form and structure. Such a change demands careful study and research over a period of time so that continuity in the meaning of the test scores will not be jeopardized.

Further, the requirements to disclose the contents of all tests which are taken by more than 5,000 people per year will create inordinate hardship on the process of providing (1) achievement tests for small numbers of people at reasonable cost, (2) alternate arrangements for those whose religious beliefs preclude their taking tests on Saturday, (3) alternate (braille/cassette) versions of tests for handicapped students, (4) additional testing opportunities on special dates for those in large urban areas, and (5) special testing for the military on military bases. Moreover, the apparent requirement to provide an individual with his/her own (as opposed to a copy of his/her) answer sheet would raise serious questions of legal evidence in the event that disagreement arose as to the correct score earned by that individual.

(5) ETS supports both the intent and the language of Section 6, although the section might benefit from inclusion of language designed to protect the confidentiality of institutional data.

(6) While ETS supports the intent of Section 6, there may be questions concerning the appropriateness or legality of requiring disclosure of financial information at the level of detail required. Of greater importance, there are several problems with definitions (e.g., the distinction between a test sponsor and an agency with which the sponsor contracts) that should be clarified in order to make the provisions of this section workable. Finally, some of the specific data required (e.g., the number of times a test was retaken during a single year) is not currently collected.

(7) Some of the definitions in Section 10 are ambiguous.

(8) The 180 day implementation period will be virtually impossible, particularly if the legislation is passed after materials for a given teting year have been printed and distributed nationally.

SOME IMPLICATIONS ON TEST DEVELOPMENT COSTS AND TEST AVAILABILITY OF NATIONAL TEST DISCLOSURE

Some of the basic problems of compliance with the provisions of H.R. 4949 as currently written are directly related to the requirement for disclosure of all test forms. The following examples - based upon the current administrative patterns of the Graduate Management Admission Test (GMAT) and Graduate Record Examinations (GRE) Advanced Tests - will demonstrate several of the more significant problems. Similar tests - the Law School Admission Test, the GRE Aptitude Test and others - will face similar problems.

Both the GMAT and the GRE Advanced Tests are administered nationally - that is at approximately 500 sites across the country — on both Saturdays and Mondays to accommodate those whose religious beliefs preclude their testing on Saturday. As the bill is currently written, six of the GRE Advanced Tests (achievement tests on specific subjects) would exceed the 5,000 limit; thus, these tests would have to be disclosed after each administration on six Saturdays and five Mondays during the year. Similarly, the GMAT would have to be disclosed after each administration: four Saturdays and four Mondays.

Currently, two new versions of the GMAT are developed each year. For the GRE Advanced Tests - which involve work by committees of faculty members drawn from the appropriate disciplines - only one new edition of each test is currently created each year. In order to accommodate service patterns for national and other administrations, older editions of the tests developed in prior years are repeated, a process that is made possible only by the fact that these older editions have not been disclosed, or have remained "secure." This repetition of older editions is also currently used as a key method for equating scores from one test administration to the next.

Costs for creating a new GRE Advanced Test in any discipline are approximately $50,000 for each new version, and the costs of each version of the GMAT is approximately $92,000. As can be seen from the Table below, the costs of creating enough additional forms of these two tests to permit the current national testing pattern using "secure" forms would approach $2,700,000 annually for the six GRE Advanced Tests only and over $500,000 for the GMAT.

	Total National Administrations	Number of New Forms Now Created Annually	Number of Additional New forms to Keep Current Pattern	Approximate Cost per New Test Form	Total Additional Cost Annually	Number of Test Takers	Current 1979-80 Test Fee
GMAT	8	2	6	$92,000	$552,000	185,000	$12.50
GRE Advanced Tests	10 Administrations for each of 6 tests = 60	1 for each or 6 total	9 for each of 6 tests = 54	$50,000	$2,700,000	70,000 for 6 Tests	$14.00

As is apparent, some modification in the national service pattern would in all probability have to be considered if the current legislation were enacted. Regrettably, although the overwhelming majority of test takers are accommodated at these National administrations, the 10 GRE or 8 GMAT administrations represent only a small proportion of the total number of administrations offered. For example, when weather or other problems force cancellation of a national test administration in any location, a free makeup administration is offered - using a different test form from that used on Saturday. During 1978-79, makeup tests were offered on 38 separate dates for the GRE and on 23 separate dates for the GMAT during 1978-79. Similarly, the GRE program offers its tests in urban areas on other than national test dates for the additional convenience of test-takers in those areas; these so-called "Special Administrations" were offered on a total of 116 different dates in 9 cities during 1978-79. When special GRE testing for handicapped students (38 dates), for military personnel on military bases (143 dates), and testing carried out on individual campuses (57 dates) are added, the true potential impact of universal test disclosure on the need for different test forms can be imagined. Because of the value of these many alternate arrangements - including testing for Sabbath observers, for the handicapped, for the military, and for residents of urban areas - we would strongly encourage reconsideration of both the language and the intent of the test disclosure section of this law so that it does not inadvertantly lead to reduction of services to any segment of the population.

Educational Testing Service

Principles, Policies and Procedural Guidelines Regarding ETS Products and Services

February 1, 1979

PREFACE

Educational Testing Service (ETS) is a private, nonprofit, educational organization with primary involvement in the areas of measurement and research. In collaboration with a wide variety of institutions and agencies, ETS provides programs, research and services for the identification and recognition of individual talents and seeks to contribute to the expansion of opportunities available to all individuals.

ETS recognizes its responsibility to the individuals and institutions that it serves directly and to the larger society that ultimately must judge the value and efficacy of its work. This responsibility is implicit in the educational character under which ETS operates and is reflected in the products and services offered by the organization since its inception in 1947. Periodic and intense self-examination has been one means by which ETS has attempted to assure its continuing commitment to the broad constituency which it serves. As a step in that continuing process, ETS has developed these Principles, Policies and Procedural Guidelines Regarding ETS Products and Services (Guidelines), first published on August 1, 1977.

The Guidelines are designed to ensure that ETS products conform to uniformly high standards with respect to seven areas of basic importance: Accountability, Confidentiality of Data, Product Accuracy and Timeliness, Research and Development, Tests and Measurement, Test Use, and Technical Assistance, Advice and Instruction. The first three sections of the Guidelines deal with issues that relate to all ETS activities: Accountability, the responsibilities of ETS to those affected by its activities; Confidentiality of Data, the rights to and limitations on access to data collected by ETS; and Product Accuracy and Timeliness, the control of quality and performance according to scheduled commitments.

The remaining sections concern issues specifically relating to ETS' main endeavors: Research and Development, Tests and Measurement, Test Use, and Technical Assistance, Advice and Instruction.

The Guidelines attempt to codify standards used in various ETS programs and services that should be considered for more general application and more formal articulation. ETS has not developed these Guidelines because of any legal requirements. They were not devised in response to any standards published by professional organizations, although ETS endorses the goals served by those efforts. These Guidelines are the result of an extensive deliberative process, which began at my direction in 1974. More than 150 members of the staff have participated in the initial development of the Guidelines and their subsequent review and revision. The Guidelines are drawn from the particular circumstances and needs at ETS and are designed to reflect its institutional objectives. Because of their origin and purposes, the Guidelines cannot generally or usefully be applied to organizations whose practices, programs or services differ from those of ETS.

The Guidelines include three types of statements, which have varying degrees of generality and significance. The principle that should underline ETS efforts in any given area is set first; policies to govern decision-making, and designed to foster more specific goals, are set forth next; and procedures relative to the conduct of specific ETS activities are last described. However, because the guidelines must apply to a broad diversity of programs and services with differing characteristics, each procedural guideline represents only one method of achieving the objectives stated in the policies.

Even the principles and policy statements might reasonably be expected to change in focus or coverage, as scientific developments or practical experience change the nature and desirability of various goals. Thus, the Guidelines are intended to encourage and not deter change and improvement. They should foster the development, exploration and use of alternative approaches that hold to high standards and serve important ETS policies.

ETS believes that the Guidelines can contribute significantly to the quality and utility of ETS products and services. The Guidelines have been distributed to executive, managerial and professional staff at ETS so that they may be applied in the course of their work at ETS. ETS does not have complete responsibility or authority, of course, to determine how the Guidelines will be implemented in ETS programs for which policy is substantially established by a sponsoring group other than ETS. ETS has, however, taken steps to encourage and assist those groups to implement the Guidelines as their activities relate to ETS. New activities ETS enters into will be held to this same expectation.

During the first year after their introduction, the Guidelines existed in provisional form. They were used in evaluating all ETS programs and services, and, where a substantial lack of compliance was found, steps were taken to achieve compliance within a reasonable time, or to revise particular provisions to reflect more accurately the diversity and practical demands of ETS operations. This process of review will continue as part of our organization-wide quality assurance effort and a report will be given to the Board of Trustees at regular intervals. We are committed to the effort to apply a set of positive criteria to our work and, taking the present document as a starting point, we will go forward with the continuing process of improving those criteria and consequently the programs and services ETS provides.

February 1, 1979

W. W. Turnbull
William W. Turnbull
President

ACCOUNTABILITY

Principle

ETS acknowledges responsibility for the effective stewardship of its resources to the New York Board of Regents which has issued its corporate charter; to the governing boards that sponsor and set policy for programs or services in which ETS products or services are used; to the individuals and committees that advise ETS with respect to appropriate policy for its programs; to the institutions and agencies that use ETS products and services; to persons who take ETS tests (and parents or guardians of minor persons), submit data for use by ETS or for distribution to others, or participate in research and development projects conducted by ETS; and to the professional associations that are concerned with educational and psychological measurement and research.

Policies

A. ETS will furnish appropriate information to those to whom it is responsible so they may make informed, independent judgments as to the effectiveness with which ETS exercises its stewardship.

B. ETS will seek, consider and, as appropriate, act on the views of those who sponsor, use or are affected by ETS programs and services.

C. ETS will seek to obtain advice on its activities and policies from qualified men and women who are not employed or retained on a regular basis by ETS and who are drawn from appropriate professional disciplines, major philosophies and points of view, different geographic regions, and the major ethnic groups within the relevant population.

D. ETS will support the activities of professional associations with respect to developing and implementing professional standards or codes, making available the results of current work, and fostering peer review of its activities.

Procedural Guidelines

1. Information should be provided to sponsoring organizations with which ETS has contractual relationships in a form that permits evaluation of ETS services in terms of:
   - a) quality;
   - b) timeliness;
   - c) costs; and
   - d) responsiveness to legitimate comments or criticisms.

2. Procedures should be established to facilitate communication with sponsors by:

   - a) meeting at least annually to provide information and to receive comments on matters affecting the operations with which they are concerned;

   - b) defining a mutually agreeable process to be used to transmit comments from sponsors or others and a time period within which the evaluation of comments will be completed and reports of actions to be taken by ETS can be expected.

   - c) making available periodic opportunities for sponsors to express opinions, judgments and counsel concerning their activities or programs directly to ETS officers not normally responsible for such activities or programs.

3. Procedures should be established for making available technical and other information about ETS products and services to users so that they may evaluate the appropriate use of the product or service and communicate comments or criticisms to ETS.

4. Procedures should be established to communicate with or provide information to persons who use or take ETS tests, who submit data for use by ETS or for distribution to others, or who participate in research and development projects conducted by ETS. This information should be communicated by ETS or the sponsor in such a way that these persons may understand their participation with respect to:

   a) the identity and scope of the sponsor's responsibility;

   b) the nature of the product, service or research by which they are affected;

   c) the way in which the product, service or research will likely be used by educational institutions or others; and

   d) the channels that have been established for addressing comments or criticisms to ETS or to the sponsor and response thereto.

5. Organizational and program financial information should be recorded, processed and reported in accordance with generally accepted accounting principles and under appropriate safeguards to insure accuracy.

6. An annual report that provides information about organizational activities and finances should be published by ETS on a regular basis and made available to any person on request. Program and project reports, including program financial information, should be made available in a manner consistent with contractual understandings.

7. Requests for information that is not included in an existing publication should be considered by the appropriate sponsor and by ETS. If its disclosure is consistent with applicable law, with ETS and sponsor policy, and with contractual obligations governing confidential or proprietary information, the information should be provided. If complying with a request for information results in a cost to ETS or a sponsor or affects the normal schedule of fulfilling ETS' responsibilities, ETS may provide the requested information in a reasonable period of time and at an appropriate price for the services rendered. Procedures should be established, as appropriate, to facilitate responses to these requests.

8. Changes in federal statutes, regulations and case law that affect research and development, testing programs, or advisory and instructional services should be monitored to insure that ETS activities and operations are in compliance as relevant federal laws or rules change. Changes in other statutes, regulations and case law should be evaluated as appropriate for the same purpose.

9. All proposed new ETS activities should be reviewed by counsel for compliance with applicable federal law and state law as appropriate. ETS officers and staff should direct the attention of legal counsel to matters that might affect ETS compliance.

10. Advice should be sought, where appropriate, from men and women drawn from diverse backgrounds, interests and experience (e.g., appropriate professional disciplines, major philosophies and points of view, various geographic regions, and major ethnic, handicapped and other relevant subgroups of the population of interest) who are qualified to make a contribution to the direction and substance of ETS programs and who are not employed or retained on a regular basis by ETS.

11. Individuals who become members of an ETS external advisory, review or evaluation committee should be informed about the results of the committee's work in a reasonable period of time.

12. A reasonable accommodation should be made with respect to the professional responsibilities of the staff in order to permit staff members to attend professional meetings, to contribute to the development of professional standards or codes, to participate in and benefit from the dissemination of information on subjects of professional interest and to stay abreast of current concerns and accomplishments in related fields.

13. ETS should have effective procedures for peer review whenever that will contribute substantially to the quality of ETS work.

14. ETS should have effective and equitable procedures for handling questions of score authenticity arising in connection with the administration of tests.

CONFIDENTIALITY OF DATA

Principle

ETS recognizes the right of individuals and institutions to privacy with regard to information supplied by and about them that may be stored in data or research files held by ETS and the concomitant responsibility to safeguard information in its files from unauthorized disclosure.

Policies

A. ETS will ask individuals to provide information about themselves only if it is potentially useful to those individuals, necessary to facilitate processing of data or serves the public interest in improving understanding of human performance. Insofar as possible, individuals should be informed of the purpose for which the information is requested.

B. The right of individuals to privacy regarding information about them that may be stored in the data or research files held by ETS extends both to processed information, such as scores based on test-item responses, and the raw data on which the processed information is based.

C. ETS will protect the confidentiality of data supplied by institutions or agencies about themselves, and so identified, to the extent that such confidentiality does not conflict with ETS obligations to individuals.

D. ETS will not collect or maintain in its data or research files any critical information that in its judgment cannot be protected adequately from improper disclosure.

E. ETS will encourage the organizations with which it works to adopt policies and procedures that adequately protect the confidentiality of the data transferred by ETS to those organizations.

Procedural Guidelines

1. Information about an individual, which has been identified as such, may not be released by ETS to organizations other than those for which the information was collected without the consent of that individual. A written exception may be made in the case of research studies during which the Committee on Prior Review of Research has determined that release of the data serves a public need, that there is no satisfactory and reasonable alternative way of obtaining the information, that the recipient researcher will use the data in appropriate ways and that there are adequate assurances of confidentiality.

2. Information about an institution, which has been identified as such, may be released from ETS only in a manner consistent with a prior agreement or with the consent of the institution or with the approval of the cognizant ETS officer and representative of the appropriate sponsor (if any).

3. An individual should be able, on payment of a reasonable fee, to authorize the disclosure of information about himself or herself from program data files held by ETS to any appropriate recipient, provided that such authorization is in writing and that disclosure is not inconsistent with other ETS or sponsor policies and does not violate the privacy of other individuals. Identification of the requester, through signature and data file number, or other appropriate method, should be required before any such information is released.

4. In an emergency and when it is to the benefit of the individual, an authorization by telegram or telephone for the release of personal data should be acceptable, provided that such authorization includes adequate identifying information and that such release is not inconsistent with other ETS or sponsor policies. By prior agreement with the individual, authorization by a designated agency or institution should also be acceptable. In such instances, the individual should be informed that the disclosure has taken place.

5. If an individual is not competent because of illness or other considerations, information about that individual may be released from data files only with the consent of the individual's parent or legally appointed guardian.

6. Unless the access to confidential data can be safeguarded, ETS should not participate in any time-sharing network, data bank, or other electronic data processing or storage system involving units outside ETS.

7. On submission of appropriate identifying information and payment of a reasonable fee, an individual should be able to obtain information about himself or herself in ETS-held data files for the following purposes: to ascertain the accuracy of personal or biographical data and to request verification, within a reasonable period of time, of test scores or other processed information from tests, questionnaires, or school records, provided such release is consistent with sponsor policies.

8. Procedures should be developed for systematically eliminating from data files information that is judged to be out of date and, hence, of minimal value.

9. Information from ETS-held program data files provided by individuals for a designated purpose should not be used or released for another purpose (such as a validity study or research project) without the individual's consent except when used or released in a form that cannot be identified with the individual.

10. ETS should refuse to provide personally identifiable information except in accordance with these guidelines unless served with a subpoena or other court order. In that event, ETS should make appropriate efforts to quash or narrow the subpoena or order or to obtain a protective order to minimize the exposure of personally identifiable information.

11. At the time information is collected and to whatever extent practical, programs should inform individuals of the conditions surrounding the release and confidentiality of the information about them.

12. Individuals should be identified in ETS research files only by code numbers. Information linking the code numbers to names should be kept in a secure location only as long as necessary for purposes such as follow-up studies or collating new data, after which the names should be destroyed.

13. Every organization with which ETS works should be informed of the confidential nature of any data transferred by ETS to that organization or collected by the organization on behalf of ETS so that appropriate procedures can be employed by the recipient organization to protect the confidentiality of such data.

PRODUCT ACCURACY AND TIMELINESS

Principle

The accuracy of ETS' principal products and the timeliness with which they are made available are important parts of the responsibility ETS has undertaken with respect to its sponsors and the diverse public it serves.

Policies

A. ETS will establish standards of accuracy and timeliness with respect to each principal product.

B. ETS will use quality controls that are adequate to assure that its standards of accuracy and timeliness are met.

C. ETS will make realistic delivery commitments and reasonable efforts to meet those commitments.

D. ETS will sacrifice the timeliness of the delivery of information if the desired accuracy of that information is substantially in question.

E. ETS will seek to inform those adversely affected if, subsequent to its release, information has been found not to meet ETS standards of accuracy.

F. ETS will seek to inform those adversely affected if there is a probability that there will be substantial departure from ETS standards of timeliness with respect to a principal product.

Procedural Guidelines

1. Principal products should be identified and a standard of accuracy using units of measurement appropriate to the type of product should be established for each.

2. When appropriate, quality control should include an adequate and independent recomputation and a visual reexamination of ETS-processed information based on an appropriate sample of cases sufficient to identify errors within the limits of the applicable standards of accuracy.

3. When the computational nature of the information is such that it is impossible or impractical to determine the accuracy of the information by independent recomputation, staff members who are technically competent to do so should assess its "reasonableness" as a part of quality control.

4. There should be a quality inspection of intermediate products when:

   a) the accuracy of variable information (e.g. parameter data, algorithms), verified by independent recomputation or assessment, influences the nature of an ETS process or computation and is critical to the process that generates an ETS principal product; or

   b) detection and correction of errors would facilitate meeting the delivery commitment on the principal product; or

   c) the nature of the principal product is such that it is impossible or impractical to determine the accuracy of the information by independent computation using the source data.

5. Quality control for principal products that do not contain processed information (such as bulletins of information or test books) should include inspection of a sample prior to release of the product. If the product is released from an outside vendor (e.g., outside publisher) or a sponsor's agent, quality control should include inspection of those components of the principal product that contain critical information on ETS-provided services.

6. Quality control of information given in letter or telephone responses should include a periodic audit of a sample.

7. Failure to meet standards of accuracy and timeliness should be reported to a designated ETS staff member for resolution.

8. A principal product that does not meet established standards of accuracy should not be released until appropriate corrective action is taken unless release would be for the benefit of the score recipient and users and permission to release is given by the cognizant ETS officer.

9. If an error is found in critical information already released by ETS, the correct information should be promptly distributed.

10. Process control methods (e.g., a predefined schedule including a delivery date and contingency procedures for dealing with volume surge) should be established for the production of each principal product to help assure its delivery by the scheduled delivery date.

11. If it is likely that there will be a substantial departure from ETS standards of timeliness with respect to a principal product, those who would be adversely affected should be so notified.

RESEARCH AND DEVELOPMENT

Principle

A continuing program of research and development conducted in compliance with professional standards with respect to quality and ethical procedures is necessary to maintain the high quality and social utility of ETS contributions to education. This includes basic inquiry to increase understanding of educational processes and human development; evaluative and applied research in response to the needs of the educational community; and research and development to improve ETS products and services. Publication of the results of significant ETS research is of benefit to ETS and the profession because it permits others to use, build upon or improve ETS work.

Policies

A. ETS will devote appropriate research efforts to improving education through the discovery and conceptual integration of new principles and understanding. This research will be aimed at extending knowledge of the learner and learning processes, of learning environments and educational treatments, of educational institutions and of the interacting factors that influence human development.

B. ETS will devote appropriate research efforts to the improvement of the technical quality of ETS products and services. Among the important issues addressed by this research will be problems of test development, reliability, equating, validity, and meaningfulness of interpretation.

C. ETS will devote appropriate research and development efforts to the identification of needs of the educational community and to the creation, improvement and evaluation of instruments, systems and programs of service that meet these needs.

D. ETS will conduct its research under appropriate procedures that protect the rights of privacy and confidentiality of human subjects or respondents.

E. ETS will follow procedures to insure that ETS research is of high quality. Standards of quality in research refer to such matters as the identification of relevant data, the choice of suitable methods of collecting and analyzing data, the logic and objectivity of analysis and interpretation, the exploration of relationships between research problems and findings, on the one hand, and existing knowledge, theories and methodologies on the other, and the thoroughness and care of project planning and management.

F. ETS will undertake research only if its potential benefits outweigh the inconveniences of or risks to the subjects or respondents who are involved.

G. ETS will encourage the dissemination of full accounts of ETS research in the usual professional forums and will provide internal means by which the results of ETS research can be published.

Procedural Guidelines

1. To maintain the quality of operational programs, ETS should engage in the following activities:

   a) study and research on the test development process, including systematic development and evaluation of new item types and approaches;

   b) studies to determine the sources of significant differential performance of sex, ethnic, handicapped, and other relevant subgroups on ETS tests;

   c) periodic evaluation of current approaches to aptitude and achievement measurement to determine fairness, validity and appropriateness for significant subgroups such as minorities and women;

   d) research related to reliability theory and practice, including methods of determining the reliability of classification decisions;

   e) study of the equating methods presently in use and development of improved methods as limitations in the applicability of the present methods are observed; and

   f) research to advance measurement techniques and selection and classification models relevant to fairness and validit

2. Research projects should be undertaken in such areas as learning and cognition, personality and social influence, teacher behavior and instructional processes, socialization and human development, and the economics and sociology of education as a means of improving educational policies and practices.

3. Efforts should be made to develop instruments and programs of service in areas such as measurement, institutional

and program assessment and evaluation, instruction, guidance, financial aid, certification and licensing, and technology that would be of educational and social utility.

4. Proposals for research to be conducted by ETS and involving human subjects or respondents should be considered by the Committee on Prior Review of Research, under its procedures for review, to verify that proper arrangements have been made for protection of the welfare and rights of human subjects.

5. Researchers should not conduct research projects without the consent of subjects and respondents. In the case of young children, the consent of parents or a legal guardian, or of appropriate institutional representatives, should be obtained.

6. Each research proposal should be reviewed by one or more persons who are competent in the field within which the proposal falls. They should be satisfied that professional standards of quality and ethical conduct are met.

7. Identifiable data should be released from ETS to researchers other than those who originally conducted the research only when one of two conditions have been met:

   a) Consent to do so has been given by or on behalf of the subjects or respondents or by those who have given consent on their behalf; or

   b) the Committee on Prior Review of Research has determined that release of the data serves a public need, that there is no satisfactory and reasonable alternative way of obtaining the information, that the recipient researcher will use the data in appropriate ways and that there are adequate assurances of confidentiality.

8. After the data-collection phase of a research project has been completed, subjects should not be expected to provide additional data for a follow-up study unless such participation was part of their original agreement to serve as subjects, or their consent for follow-up is obtained or the follow-up study has been approved by the ETS Committee on Prior Review of Research.

9. The results of measures of performance based on experimental situations or tests the interpretation of which is therefore tentative and whose applied use is not yet supportable should not be reported to subjects, or to the institutions providing the subjects, unless there is relatively little danger of misinterpretation or misuse of the information that would be harmful to those individuals or institutions or unless the use is part of a feasibility study or experimental condition. Stipulations regarding nonissuance of such reports should be made to participants in advance of the data collection.

10. The results of each research project undertaken with respect to a particular ETS program or service should be available for dissemination unless a specific need to restrict publication to protect confidentiality or for other program purposes is identified prior to the beginning of the project and made known to the appropriate individuals.

11. The contracts under which research is undertaken for agencies or institutions outside ETS should permit publication of the results of the research unless a specific need to protect the research results is identified prior to the beginning of the research and made known to the appropriate individuals.

TESTS AND MEASUREMENT

This section which deals with ETS testing activities is divided into seven subsections that are devoted to test development, test administration, reliability, scale definition, equating, score interpretation, and validity.

TECHNICAL QUALITY OF TESTS

Principle

> High standards of quality and fairness in constructing, administering, reporting, interpreting and evaluating ETS tests are central to ETS' capability to function effectively as an educational service and research organization.

Policies

A. ETS will strive to develop tests in which the attributes measured, procedures followed, and criteria used will be unbiased with regard to a heterogeneous group of examinees and appropriate to the use for which the test is designed.

B. ETS will establish standards for test-administration processes that minimize variations in test performance due to circumstances or conditions not relevant to the attributes being measured.

C. ETS will establish for its tests a high degree of reliability (accuracy of measurement), consistent with the requirements and the purposes of the test.

D. ETS will develop scales for reporting scores in a rational fashion, consistent with the requirements and the purposes of the test.

E. ETS will provide equating systems, when appropriate, for the perpetuation of scales for reporting scores at the highest level of precision practicable.

F. ETS will make available to sponsors, institutional or agency users and examinees data for interpreting scores on ETS tests that foster appropriate use of those scores.

G. Recognizing that test validation is a responsibility of both test users and test developers, ETS will encourage and assist test users in their validation efforts and will itself make available tests that are designed to meet professionally acceptable standards of validity provided the use of such tests is consistent with the primary purposes for which the tests were developed.

H. ETS will adhere to appropriate professional standards such as those published in *Standards of Educational and Psychological Tests* and *Principles for the Validation and Use of Personnel Selection Procedures*.

Procedural Guidelines

Section 1: Test Development

1. Policy and substantive contributions to the test development process should be obtained from qualified men and women who are not on the full-time staff of ETS and who are drawn from diverse backgrounds and appropriate specialties within professional fields (e.g., various kinds of institutions and programs, relevant philosophies and points of view, and major ethnic, handicapped and other relevant subgroups of the population).

2. Appropriate background information for use in the development of a test should be documented at appropriate stages in the development process and include:

   a) the purpose for which the test is intended to be used;

   b) the nature of the population that will take the test;

   c) the relevant procedural, financial or time constraints that will influence the available test development methods and their likely outcomes;

   d) for achievement tests, the kinds of curricula for which the test is designed;

   e) for job-related tests, the elements in training or employment that are related to performance on the job.

3. For each test, specifications should be developed and reviewed by a process that provides information from the following perspectives:

a) content and skills--specifications should include the psychological, educational, or other domains to be sampled; the relative weight to be given to each domain; the appropriate level of proficiency to be required within each domain; a balance with respect to curricular differences.

b) test and item format--specifications should include the item (question) types that are most clearly related to content or skills to be measured; the appropriate level of language or reading; requirements regarding clear and comprehensive directions and sample items or the need for a sample test; and whether free-response, multiple-choice or other machine scorable formats can be used.

c) psychometric--specifications should include the level of difficulty of the test; the distribution of item difficulties (when pretested items are used); guidelines for evaluating the homogeneity among items within a test and the relationship between subtests or tests; equating requirements; number of items and time allotted.

d) sensitivity--specifications for tests should require material reflecting the cultural background and contributions of women, minorities, and other subgroups; specifications should also require a balance of positive connotations if negative connotations are made in any references to these groups.

4. Except for tests designed to measure rate of performance, the number of items in a test that has a specified time limit should be chosen so that time is not a decisive factor in performance, at least for the large majority of examinees.

5. Subject matter and measurement specialists familiar with the purpose of the test and with the characteristics of the intended population should review the test items for accuracy, content appropriateness and the adequacy with which the items sample the domain.

6. The individual items in a test should meet appropriate technical standards such as those contained in the manuals for item writers used in the test development area.

7. Individual test items and the test as a whole should be reviewed to eliminate language, symbols or content which are generally considered potentially offensive, inappropriate for major subgroups of the test-taking population or serving to perpetuate any negative attitude which may be conveyed toward these subgroups. No item in any test should include words, phrases or description that is generally regarded as biased, sexist or racist (e.g., demeaning modifiers and stereotypes).

8. The items in a test should be reviewed by editorial specialists for clarity, accuracy, consistency, and, when appropriate, for conformity with standard editorial style.

9. Tests should contain clear and complete directions. Enough sample problems should be provided in test-program publications so that the examinee can understand the nature of the task and the test-taking procedures. Where there is a need to provide a general orientation to testing, as when testing young children, practice tests--included either in descriptive material or at the time of test administration--should be used.

10. The typography, directions, and arrangement of items in the test booklet should facilitate the task of test-takers. When appropriate, tests should be made available to handicapped individuals such as sight-deficient candidates through tapes, readers or special printing.

11. Methods should be employed to evaluate the appropriateness of items before their operational use in a program or before the reporting of scores. Appropriate methods include pretesting, preliminary item analysis (using the first operational use of items as an opportunity to identify inadequate items) or careful review of the results of administering similar items to a similar population. In assessing the appropriateness of items before their operational use, efforts should be made to include representative samples of the operational test-taking population.

12. The operational use of each test should be followed by systematic item analyses using appropriate criteria and by test analyses. These analyses should include reliability, intercorrelations of sections or parts, and speededness.

13. Studies relating item performance to subgroups should be carried out for new or substantially revised tests when there are adequate data concerning sufficient samples of large subgroups whose education and experience may be different from the majority of examinees.

14. The specifications for tests in ongoing programs should be reviewed for relevance and appropriateness before each new form is created. ETS staff and advisers should consider whether changes in the field, discipline or curricula require a revision of the specifications.

15. When major changes are made in test specifications, consideration should be given to the implications of such changes for score comparability and whether it is necessary to change the test name or otherwise communicate to those who interpret test scores that comparisons with earlier tests may be inappropriate.

16. When test forms are used for a number of years in a program, they should be reviewed periodically for their appropriateness. The frequency of such review should be determined by the amount of change occurring in the population of test-takers or the subject matter domain. Test forms that are found to be outdated should be revised or withdrawn from use.

Procedural Guidelines

Section 2: Test Administration

1. Information should be made available to prospective examinees and (in some programs, to parents or guardians as well) in advance of the test administration with respect to the following, as appropriate:

   a) the purpose of the test and what it measures;

   b) the nature of the test items (including samples of typical item types);

   c) the relevant instructions for taking the test, including instructions for guessing, changing answers, and strategy involving speed and accuracy in taking the test;

   d) identification requirements and the consequences of not having identification;

   e) the consequences of misconduct by the test-taker;

   f) background and experience relevant to test performance;

   g) the location of test centers, the test dates and special testing arrangements that can be made;

   h) the procedures for registering for the test and changing the centers;

   i) the structure of test fees and fee waivers;

   j) special arrangements available for administering tests to handicapped individuals;

k) the reporting of scores;

l) procedures for canceling test scores by the candidate and reasons why ETS or the sponsor of the test might cancel scores; and

m) the procedures for registering complaints.

2. Program publications should be reviewed for language or descriptions generally regarded as biased and offensive. For example, the exclusive use of masculine pronouns should be avoided as should the implication that all persons in a given category (for instance, examinees, supervisors, counselors, or teachers) are either females or males (unless, of course, the category is logically restricted to members of a single sex). Illustrations, examples and practice items in test-information publications should represent males, females, minority and majority groups, and individuals in ways that indicate respect and awareness of valuable contributions.

3. The facilities at which tests are administered should be places that are convenient for the majority of examinees, nonsegregated and comfortable. At least portions of those facilities should be accessible to and responsive to the needs of handicapped individuals.

4. ETS should enlist test-center supervisors and staff with demonstrated sensitivity to the anticipated sex and ethnic composition of the examinee group, based on prior experience. When appropriate, persons affiliated with institutions attended by significant numbers of those examinees should be included. Minority-group supervisors and/or proctors should be employed, and test sites should be located in minority communities whenever appropriate and feasible.

5. Test-center supervisors and staff should be familiar with the procedures for administering a standardized test and should be provided with a description of the testing program, a description of the candidate population, and specific instructions for administering the test. Instructions concern such subjects as the duties of test supervisors, associate supervisors, and proctors; the receipt, storage and return of test supplies; the admittance of examinees to the testing rooms; the distribution of test materials; procedures to be followed in administering tests to handicapped individuals; procedures to be followed in instances of suspected cheating; procedures to be followed in other cases of candidate misconduct; and procedures to be followed in case of emergency.

6. Test performance can be affected by the psychological atmosphere of the testing center. Test supervisors should be informed of this and instructed to take measures to avoid an adverse situation. For example, test supervisors should be instructed, when it is appropriate and feasible, to have minority- as well as majority-group persons, women as well as men, read test directions and to recognize questions from examinees following an impartial procedure.

7. ETS should provide the test center supervisor with directions to be read aloud to examinees before the test begins. These directions should include information relating to: procedures for marking answer sheets, timing of test sections, strategies for guessing, time and duration of test breaks and examinees' use of unauthorized aids. Test supervisors should check to see that examinees understand their task and the procedures to be followed.

8. Reasonable efforts should be made to eliminate opportunities for examinees to attain scores by fraudulent means by stipulating requirements for identification, assigning examinees to seats and requiring appropriate space between seats.

9. Appropriate procedures should be applied after the test administration to identify scores of questionable authenticity, to resolve issues of authenticity and to provide for prompt reporting of questioned scores found to be authentic.

10. A systematic program for observing test administrations should be conducted by trained ETS staff members or other qualified individuals: to review the testing procedures with the test supervisors, to insure appropriate testing conditions, to insure adequate maintenance of test security at the test centers and to relay questions and concerns from the field to the appropriate ETS office.

11. Testing programs should have detailed procedures for investigating and resolving examinees' complaints of irregular test administration or score reporting.

12. Comments and suggestions should be solicited from supervisors by such means as the Supervisor's Comment Sheet and meetings of supervisors to provide ETS staff with information to improve future administrations.

13. Supervisors should be required to record and report to ETS information on irregularities (such as mistiming, defective materials, power failures and cheating) so that ETS can evaluate the possible effect of such occurrences on examinees' test performance.

14. An individual who has taken a test should be provided information that will be helpful in interpreting scores on that test.

Section 3: Test Reliability

1. When test scores are reported to institutional or agency users or to individual examinees, information about the reliability of the test should be documented and should include:

   a) a reliability coefficient and an overall standard error of measurement (several indices may be provided if more than one method of assessing reliability has been used; alternate-form information should always be provided if available);

   b) standard errors of measurement for score regions if decisions about individuals are made in those score regions and if the overall regions and the overall standard error are judged inappropriate;

   c) the formula(s) used to estimate reliability and/or appropriate references;

   d) a justification of the method(s) used to assess reliability;

   e) a specification of the major sources of measurement error accounted for in the reliability analysis;

   f) a specification of the time interval between testings if alternate-form or test-retest reliability is used;

   g) the number of observations, the mean and standard deviation of the analysis sample (ranges or averages are acceptable in cases where the reliability information is derived from several samples);

   h) speededness data; and

   i) correlations of subscores within the same test or battery of which the test is a part.

2. If reporting any of the reliability information required under Guideline 1 is inappropriate, the reasons should be stated in appropriate program documents and, if possible, alternate information about consistency should be provided.

3. Efforts should be made to provide reliability information in an appropriate form to the examinees to whom the scores are reported.

4. The method(s) used for assessing reliability should:

    a) take into account the most common sources of error generally considered significant for test interpretation (e.g., guessing, instability over time, item and content variation, and rater inconsistency): and

    b) be appropriate to the nature of the test, in order not to seriously over- or underestimate reliability.

Procedural Guidelines

Section 4: Scale Definition

1. Raw scores on a test or subtest (including percentages of questions answered correctly) should not be reported by ETS for individual examinees or in summary form for groups of examinees except under either of the following circumstances:

    ✓ a) when it is anticipated that only one edition of the test will be offered for use in the foreseeable future or it is demonstrated by appropriate empirical procedures that raw scores on all the editions to be compared are interchangeable; or when raw scores on that test edition will not be compared directly with raw scores on another test edition; or

    b) when reported in conjunction with a scaled score and in a context that supports appropriate interpretation, such as when a copy of the test itself is available or when individual or group responses to individual items, depending on whether individual or group performance is being assessed, are available.

2. If a test or test battery yields multiple scores for an individual and scaled scores are to be used directly (i.e., without reference to norms tables) in interpreting performance profiles, the scales should be normatively defined and each should be defined with respect to the same population.

3. When different tests in a program are taken by different examinees whose scores are to be directly compared, the scales for the tests should take into account possible differences among the groups of examinees who take the various tests.

4. Established scales should not be redefined except under compelling circumstances. If a scale is to be substantially redefined, the numerical values should be changed substantially to minimize the possibility of confusion between test results expressed on the revised scale and results expressed on the original scale. An exception to this guideline may appropriately occur if the test in question is one of a set of tests for which a single range of numerical values (e.g., 20-80) is used and the scales for other tests in the set have not been redefined.

5. Scale properties that affect score interpretation and use should be described in program publications available to the examinees and to institutional or agency users.

6. Technical manuals and interpretive publications for institutional or agency score users and examinees should indicate, in language appropriate to the audience, whether a distributively based scale is intended to be normative or nonnormative. If it is intended to be normative, the group should be described.

7. Whenever a normatively defined scale no longer conveys useful normative information, all published descriptions of the scale should be changed accordingly.

8. Program publications should caution score recipients (users and examinees) that scores received on different tests that are reported on scales that are similar in appearance may not be equivalent.

Guidelines 9 through 14 apply only to scales established after guidelines 1-8 were published on August 1, 1977.

9. If a scale is to be distributive, the choice between a normative and nonnormative distributive scale should take into account:

   a) the extent to which normative interpretation with reference to a particular population will be appropriate and useful for all examinees who take the test and for all purposes for which the scores are intended to be used;

   b) the probable time period during which the normative information conveyed by the scores will continue to be descriptively appropriate; and

   c) the feasibility of identifying and testing a suitable group of examinees on which to base a normative scale.

10. The choice between a distributive and nondistributive scale should take into account the use for which the test was intended and to which the test is likely to be put.

11. If a scale is to be defined with reference to standards of performance, the basis for establishing the standards should be determined empirically or rationally rather than arbitrarily.

12. The conventional grade- or age-equivalent score (the grade or age for which a particular score is the average) should not be used to establish the score scale for a test or system of tests. This type of score, as it typically has been derived, should be avoided altogether as a basis for reporting test performance. However, the grade (or age) for which a particular scaled score on a test is the average, referred to here as a "grade (age) level indicator" to distinguish it from the conventional grade-equivalent (age-equivalent) score, may be reported to help in score interpretation, if the practices customarily followed in deriving and presenting grade-equivalent (age-equivalent) scores are modified in accordance with ETS criteria that obviate the technical interpretive problems that grade-equivalent (age-equivalent) scores create.

13. The choice of a scale should take into account the likelihood of confusion with other widely used scales.

14. In establishing the number of distinct scale values to be reported, consideration should be given to the relative importance of the need to avoid erroneous distinctions among individuals (by reporting different scores for individuals whose true scores are the same) and the need to maintain distinctions that, on the average, will be correct (by reporting different scores for groups of individuals whose average true scores are different).

Procedural Guidelines

Section 5: Equating

1. Adequate equating should precede comparisons of the test performance of two or more individuals or groups on nonidentical items or sets of items such as test offerings in which successive, or alternate, forms are used interchangeably.

2. Statistical methods selected for equating should be used only under circumstances that are consistent with the assumptions under which the methods have been developed.

3. In regular and continuing testing programs that are available to users, integrated, long-range systems of equating the scores to all successive editions of the test should be used and described in technical publications.

4. For those tests that are offered for institutional use (as distinguished from externally administered tests offered in testing programs) of which only a limited number of forms are available, equating of new forms should be based on specially designed studies in which examinees or groups of examinees are selected by an appropriate sampling procedure to take the alternate forms or alternate sequences of forms.

5. When test forms are equated with the use of common (anchor) items, the psychological task of taking those items (represented, for example, by the directions, the context of the items and the speededness of the part of the test in which the items appear) should be the same for all examinees.

6. When the common items used for equating are not representative of the tests being equated, the groups of examinees used for equating should be as nearly as possible equivalent.

7. In the continuing testing programs, statistical checks (e.g., check equating, special scale-stability studies) should be employed to permit regular assessment of the precision of the equating.

Procedural Guidelines

Section 6: Score Interpretation

1. Effective test use and meaningful score interpretation should be supported and augmented by:

   a) the development of appropriate test norms based on administering tests to samples from a defined population when there is a reasonable expectation that a large proportion of the schools or other units selected for the norms sample will agree to participate; or,

   b) a rationally developed system of interpretation shared with score recipients when score interpretation is not developed from normative data.

2. Tests offered for sale and described by ETS as standardized tests (as distinguished from tests offered in testing programs) should have adequate norms or other information for use in interpreting test results.

3. When test norms are developed by administering tests to samples from a defined population, the resulting norms should be representative of any relevant subgroup, including those defined by sex or ethnicity, in proportion to their frequency in the defined population. Such subgroups may be deliberately over-sampled for more precise estimation of the statistical characteristics of the population by procedures that take over-sampling into account. Data on the proportions in the sample and in the population, when available, should be reported in an appropriate technical publication.

4. The report of a special norms study should provide information on:

    a) the sampling design;

    b) the participation rate of institutions or individual respondents in the sample;

    c) characteristics of the participating institutions and individuals;

    d) weighting systems used in preparing norms; and

    e) estimates of sampling variability along with an acknowledgment, when necessary, that such estimates do not take into account biases arising from nonparticipation.

5. When descriptive statistics based on program testing (as distinguished from norms based on special norms studies) are published, the following guidelines should be used:

    a) both table titles and descriptive material should make it clear that the statistics are based on examinees or participating institutions or other using agencies;

    b). the descriptive material should define the nature of the group by identifying the appropriateness of the sample and the factors that relate the background of the group to test performance, and by acknowledging explicitly that the sample is self-selected;

    c) when possible, reports should be prepared to show comparisons of data based on program examinees or institutional characteristics with relevant data on variables from other sources;

d) when information about interpretive data is prepared for different user groups, the presentation, whenever practicable, should be adapted to the needs and background of each group.

6. When norms are developed from program testing, the age, sex and ethnic composition of the program norms group should be described whenever such information about subgroup membership is available.

7. In testing programs, descriptive statistics should be compiled periodically from a sample or entire population in order to monitor the participation and performance of males and females drawn from diverse backgrounds, interests and experience (e.g., major ethnic group, handicapped status and other relevant subgroups of the population of interest).

8. If norms intended for use in the interpretation of individual scores are presented separately for males and females or for members of specific ethnic groups, the rationale should be carefully described. Separate norms may be justified for scores used primarily for guidance when access to the experiences needed to earn a high score is clearly related to subgroup membership and a more direct index of access is not available. The existence of score differences between subgroups does not in itself justify presentation of separate norms.

9. Descriptive statistics prepared separately for subgroups of the relevant test-taking population but not intended for use in interpreting individual scores should not be presented in a way that encourages their use for such a purpose.

10. Institutional or agency users and examinees should be informed of the standard error of measurement of a score, and test interpretation materials should point out the limitations of test scores and encourage score users to take into account the possible scores a test taker might achieve on retesting.

11. Statistical data used in score interpretation should be revised annually except when less frequent revision is judged to be appropriate as, for example, when norms are based on special studies. A statement of the period in which the data were collected should be included in any publication that presents the data.

13. Institutional or agency score recipients should be provided with interpretive materials designed to be helpful for using scores in conjunction with other information, setting cutting scores where appropriate, interpreting the scores for special subgroups (e.g., ethnic minorities, males, females, and handicapped students), conducting local normative studies, and developing local interpretative materials.

Procedural Guidelines

Section 7: Test Validity

1. ETS should provide evidence of the validity of its tests in relation to the principal purposes or intended uses of the tests. One or more of the following may be applicable:

   a) when test scores are to be interpreted in terms of degree of mastery of the knowledge, skills, or abilities of a domain represented by the test, content validation evidence should be provided.

   b) when test scores are to be interpreted in terms of the prediction of future behavior, criterion-related validation evidence should be provided.

   c) when test scores are to be interpreted as a measure of a theoretical construct, construct validation evidence should be provided.

2. Evidence of content validity should be based (a) on a careful determination and analysis of the domain(s) of interest and of the relative importance of topics within the domain, and (b) on a demonstration that the test is an appropriate sample of the knowledge or behavior in the domain(s). A report on evidence of content validity should present descriptions of the procedures employed in the study, including the number and qualifications of experts involved in the analysis of the domain or evaluation of the relevance and appropriateness of the test.

3. Construct validation should be based on: rational and empirical analyses of processes underlying performance on the test in question including, where appropriate, noncognitive as well as cognitive functions. Empirical evidence relevant to the analyses should include results of investigations of the degree to which test scores are related or unrelated to other variables in ways implied by intended interpretations.

4. Criterion-related validation should be used only when technically sound and relevant criteria are available or can be developed and when other conditions affecting feasibility warrant the study.

    a) Criterion-related validation should involve as many performance variables as necessary to permit evaluation of the effectiveness of test scores for predicting the types of behavior they are intended to measure.

    b) Criterion-related validation should not combine variables to form a single criterion measure unless such a procedure is justified by logical considerations or empirical evidence or the practical requirements of the intended use of the results.

    c) Criterion data should be collected in a way that permits an assessment of the reliability of each criterion variable, but with the understanding that there may be several sources of irrelevant variation, (sampling of criterion content, source of criterion ratings or data, and so forth).

5. Interpretations of correlations between test scores and criterion variables should take into account such factors as sample size, criterion reliability, possible restriction in the range of scores obtained in the validity study sample, and other contextual factors.

6. The method(s) by which any validation is accomplished should be fully documented; such documentation should include appropriate details such as the nature and reliability of the criteria, a description of the subjects used, the materials surveyed and the qualifications of the experts who made judgments regarding the appropriateness and importance of test content.

7. Where adequate methods are employed to insure equivalence of scores on alternate forms, it is not necessary that each new form be validated. New validation studies should be made if revised tests have substantial changes, such as different item types, or if they sample a revised performance domain.

8. When appropriate and feasible, the validity of a test should be investigated separately for subsamples of the test-taking population.

9. When a name of a test is established, it should not imply more than is justified by evidence of validity.

10. Information should be made available to institutional and agency users that would be of assistance to them in planning and conducting local validity studies.

TEST USE

Principle

Proper and fair use of ETS tests is essential to the social utility and professional acceptance of ETS work.

Policies

A. ETS will set forth clearly to sponsors, institutional or agency users, and examinees the principles of proper use of tests and interpretation of test results.

B. ETS will establish procedures by which fair and appropriate test use can be promoted and misuse can be discouraged or eliminated.

Procedural Guidelines

1. Program publications should:

    a) describe appropriate uses and caution against potential misuses of program tests;

    b) explain clearly that test scores reflect past opportunity to learn and discourage test interpretations that go beyond reasonable inferences from test performance;

    c) emphasize that an individual's test score should be interpreted in the context of other information about him or her;

    d) provide appropriate information about test content, difficulty, and purpose to help the institutional or agency user select instruments that meet the measurement requirements of the situation and avoid selecting, requiring or using inappropriate tests;

e) invite institutional or agency users to consult with the program sponsor and/or ETS about their current or intended uses of ETS-developed tests and identify the offices to be contacted for this purpose;

f) summarize results of research relevant to the use of the test or cite references in which such results are reported;

g) describe adequately and clearly scale properties that affect score interpretation and use;

h) advise institutional or agency users that decisions about the application of single or multiple prediction equations, based on distinguishing characteristics such as sex, ethnic group or curricular emphasis or training, should be preceded by careful examination of social, educational and psychometric factors;

i) advise institutional or agency users that if examinee grouping based on test scores is practiced, provision should be made for frequent review of group assignments to determine actual performance;

j) stress that pass-fail or cut-off scores established for such purposes as admission, credit, or certification, should be used as a basis for decision making only if the institutional or agency user has a carefully developed rationale, justification, or explanation of the cutting score that is adopted; and

k) encourage institutional or agency users to reexamine cut-off score policies periodically to minimize or eliminate possible disproportionate exclusion of members of any group such as men and women drawn from diverse backgrounds (e.g., major ethnic, handicapped and other subgroups of the population of interest) in the face of other evidence that would predict their success or indicate their competence.

2. Special (nonprogram) publications should be developed and disseminated by ETS to promote fair use of tests and discourage misuse of tests.

3. Complaints or information about questionable interpretation or use of reported scores should be investigated by means of procedures designed for detecting misuse. Such procedures should be documented, and records should be kept of such complaints and their disposition.

4. In cases where a clear misuse is brought to its attention, ETS should inform the sponsor and the institutional or agency user of ETS' opinion as to the misuse and seek voluntary correction of the misuse. If reasonable efforts to seek voluntary correction are not successful, ETS, in conjunction with the sponsor, should take steps to determine whether to continue supplying tests or reporting scores to the institutional or agency user.

TECHNICAL ASSISTANCE, ADVICE, AND INSTRUCTION

Principle

ETS is dedicated not only to providing measurement programs and conducting research but also to promoting increased understanding of measurement and test use.

Policies

A. ETS will develop and offer instructional programs in the areas of measurement, evaluation, and related research through such forms as publications, seminars, in-service training, intensive residence courses, workshops, internships and conferences. ETS may undertake these activities independently or in cooperation with other agencies, professional groups or educational institutions.

B. ETS will provide advice and information on measurement-related issues and about ETS programs, research and services. In this activity, ETS will work, where feasible, in collaboration with other professional organizations that show a concern about measurement.

C. ETS will respond promptly to requests for advice, instruction and technical assistance related both to programs and services offered by ETS and to the related areas of educational measurement, evaluation and research.

D. ETS will conform to high standards of accuracy and professionalism in its advisory, instructional and technical assistance activities.

E. ETS will provide advice, instruction and technical assistance to clients from the private and public sectors and from foreign and domestic government agencies to the extent that such services are consistent with ETS areas of expertise, meet accepted professional and ethical standards, and reflect an understanding of and respect for cultural differences.

F. ETS will endeavor to promote increased understanding of the purposes and procedures of testing among professional groups and in the public sector; ETS will make this effort both independently and in cooperation with other organizations that share this responsibility.

Procedural Guidelines

1. ETS's offices should offer advice, instruction and technical assistance; the staffing for such services should be determined by the nature of the services and the expertise required.

2. The special requirements of audiences with varying needs, interests, cultural backgrounds and levels of knowledge should be considered when ETS provides technical assistance, advice, or instruction.

3. New developments in research or testing should be considered when technical assistance, advice and instruction are offered.

4. Technical assistance, advice and instruction offered to institutions or agencies should include guidance on how to use other information about examinees (such as previous academic performance, English as a second language, and family or cultural background factors) in conjunction with test scores.

5. Comprehensive collections of reference materials relating to tests, measurement, evaluation and related research should be developed, maintained and made available to all ETS staff members and, when appropriate, to professional groups and individuals outside the organization.

GLOSSARY OF TERMS

Accuracy: The extent to which a principal product conforms to its specifications or correctly reflects the source data within the specified limits of reliability.

Client: (See Sponsor)

Consent: Permission granted by an individual or that individual's parent or guardian to the use or release of data held by ETS; such permission granted upon receipt of a reasonable explanation of the purpose of the use or release and a reasonable explanation of the manner in which the results will be reported.

Critical Information: Information that will be used to draw important inferences (a) about the sponsor, ETS-appointed external committees, institutional or agency user, examinee, subject or respondent, or (b) by the sponsor, institutional or agency user, examinee, subject or respondent and which, if incorrect, could be harmful.

Distributive Scale: A scale that is defined to yield either a specified score distribution or a specified mean and standard deviation for a particular group of examinees.

ETS Board of Trustees: The ETS Board of Trustees is the governing body of ETS. There are 16 trustees. Thirteen are elected for four-year terms. New members of the Board are elected by current trustees. Some are chosen from nominees proposed by the American Council on Education and the College Entrance Examination Board, two of the founding organizations of ETS. The presidents of the American Council on Education, the College Entrance Examination Board and ETS also serve as trustees.

ETS-held Program Data Files: Information about individuals and institutions held by ETS and derived from ETS-provided services of collection, processing, storage, retrieval and dissemination.

ETS-held Research Files: Information held by ETS and generated through ETS-conducted research intended to result in the development of new or improved techniques and materials for application in such areas as classroom instruction, evaluation of progress toward educational goals, counseling of students, and decision-making of school administrators.

Examinee: An individual who takes a test, developed and or administered by ETS.

Institutional or Agency User: An organizational recipient of ETS-processed or produced information.

Intermediate Product: Materials that are not released externally, but that are necessary to the production of the principal product.

Glossary (continued)

Nondistributive Scale: A scale that is defined without reference to the observed test performance of a particular group.

Nonnormative Scale: A scale that is based on the performance of any conveniently available subgroups of examinees for whom the test is appropriate. A score on a nonnormative scale is not intended to convey information about an examinee's standing in relation to a defined population.

Normative Scale: A scale that is based on the test performance of a sample of examinees, selected as prescribed by a specified design, from a clearly defined population. A score on a normative scale is intended to convey useful information about the performance of a particular examinee in relation to the performance of that population.

Principal Product: ETS-produced or processed materials (e.g., annual reports, performance data, score reports and admissions tickets) that are released or transmitted to a sponsor, ETS-appointed external committee, institutional or agency user, examinee, subject or respondent, pursuant to a contract or published commitment. Standards with respect to accuracy and timeliness are applicable to principal products.

Principles for the Validation and Use of Personnel Selection Procedures, Division of Industrial-Organizational Psychology, American Psychological Association. Dayton, Ohio: The Industrial-Organizational Psychologist, 1975.

Respondent: An individual who provides data to a research project in a manner and for a purpose different from either examinees or subjects.

Sponsor: Educational, professional or occupational associations, federal, state or local agencies, public or private foundations which contract with ETS for its services. This category includes their governing boards, membership, and appointed committees or staff.

Standards for Educational and Psychological Tests, American Psychological Association (APA), American Educational Research Association, and National Council on Measurement in Education. Washington, D.C.: APA, 1974.

Subgroup: A part of the larger population which is definable according to various criteria as appropriate, e.g., by (a) sex, (b) race or ethnic origin, (c) training or formal preparation, (d) geographic location, (e) income level, (f) handicap, (g) age.

Glossary (continued)

Subject: an individual who participates in an ETS laboratory or experimental research project.

Testing Program: A set of arrangements under which examinees are scheduled to take a test under standardized conditions, the tests are supplied with instructions for giving and taking them, and arrangements are made for scoring the tests, reporting the scores, and providing interpretative information as part of a comprehensive ongoing service. A program is characterized by its continuing character and by the inclusiveness of the services provided.

Timeliness: The degree to which a principal product is released or delivered to its recipient within a predefined schedule.

Mr. **Weiss.** I appreciate that all of you have busy schedules and you have already devoted the better part of the day to this hearing. We will try to proceed as expeditiously as the floor action will permit.

Let me ask a somewhat general question. Rather than to go into discussion of the philosophy which I think, given the nature of the subject, is certainly in order, I would like to start out by asking if, on the basis of the testimony, there is any factual reason why the requirements set forth in 4949, and almost equivalent information set forth in Mr. Gibbons' bill, as to whether the preregistration can be included in the legislation.

Does that present any insurmountable problem to anybody?

Mr. **Hargadon.** I think I indicated that with one exception, most of that information is already provided. I suppose this may have been Mr. Gibbons' bill which commands release of the test score that gets an individual into a given school.

Mr. **Weiss.** We had tried to undertake on the basis of the information we had, an analysis of which of the regulations under section 3 of 4949 are, in fact, included in which of the manuals. We find that although most of those requirements are included in some of the manuals, I do not think any of them really are included in all of them in totality.

For example, we found the requirement of how background information on students who have reported seems to be set forth in the LSAT manual and the GRE manual, partly, but not in the others. The score will be reported—that is set forth in the LSAT, but not in the others.

So really, although the information is available to some extent on some of the tests, the purpose of the legislative requirement would be to provide some uniformity of the information, so that even those tests which do not now include them, but which would have the information available, would also be required to submit information which certainly as far as other tests are concerned, have not been found to be onerous requirements.

Dr. **Cooper.** I do not understand your reference as to what is not reported.

Mr. **Weiss.** Section 3 of 4949 requires registration information. The testimony we have gotten from almost everyone is that most of that information in some way or another is reported or listed on the registration forms. But our analysis shows that it appears on some registration forms and not on others. The prime purpose is not to indicate, at least on my part, that none of that information is made available any place, but really to provide for a uniform requirement for that information to be set forth.

Dr. **Cooper.** In the MCAT test, all the information called for in your bill is provided either in the material sent to the individual who is going to be tested or in the student MCAT manual.

I do not think there is a single item not included.

Mr. **Weiss.** Mrs. Rice?

Ms. **Rice.** Insofar as we can understand the language, but there are some ambiguities.

Now, as stated in our testimony, we have provided all this information. I might mention some of the ambiguities in sections 3 and 4, which requests the effects of and uses of test scores.

We do not always have that information. It resides with the institution that uses the test. We advise institutions how to use our materials and tests in great detail. I am not sure exactly what you mean by "clarifying uses."

Also, under B(4) you mention major income groups. We are not sure whether you mean parental income, student income, or just what kind of financial breakdowns you have in mind. I list those simply to say that we conform to all those requirements that we can understand.

Mr. **Weiss**. Again, if I am reading it correctly as to that section of the legislation, what you are saying is, not for the most part, but the exception you have just cited, not that you cannot provide the information or that it is onerous to provide, but in fact you are saying in most instances, most of that information is already available. Your disagreement is really a policy disagreement as to whether it is necessary or advisable to make it necessary that be set forth?

Ms. **Rice**. I believe, and Mr. Hargadon probably referred to this better than I, in his testimony, there is no need for legislation in areas where there is already voluntary compliance. There are policy differences, indeed. Mr. Hargadon's testimony pointed out that we had much more trouble with the ensuing sections than with section 3.

Mr. **Hargadon**. I might add, Mr. Weiss, I am familiar with undergraduate admissions, and it is probably worth mentioning that not only does the use of tests vary widely among institutions, but within the same university, between departments at the graduate level, and other departments. In this bill, section 3, paragraph 3, the correlation is sought between test scores and future success in a career. I think the answer to that is to find out who the student intends to work for. I would think what people determine as success in the career one follows is very important at best, and I think the professional schools would be hard put ahead of time, to draw a correlation between the test score and whatever people might agree on is success in a given career.

I really do not see how we could meet that requirement.

Mr. **Weiss**. I think that is to a great extent correct. On the basis of our research, we found that there appeared to be promises made in some of the manuals which suggested exactly that kind of correlation. What we are saying, in essence is, if you are going to have that kind of suggestion or promise or half-promise of correlation between academic success or career success, then you ought to be required to set forth the basis for making that kind of promise or projection, on the basis of past results.

Would any of you have any problem with that?

Dr. **Cooper**. Mr. Weiss, as we tried to point out, admissions committees for medical schools use multivariant approaches to evaluation of a candidate. We do not want the only criterion to be any one of the criteria that are used. We think this would be disadvantageous to the applicant group.

In medicine, what is success in medicine? Is that being on Park Avenue? Is it working in one of the community hospitals? Is it working in one of the State hospitals? What is the success in medicine?

We have an extremely low attrition rate. Most of it is not due to academic failure but change in career plans. If you count success until the students finish medical school, we can say, all those accepted by the multicriteria with rare exception, those attending from 1½ to 4 years do finish medical schools and do go on to residencies and do go on to practice. I do not know how you would otherwise measure success of a physician.

Dr. **Haney**. Could I comment briefly on that?

I did not have the impression that anyone was suggesting that there is any one indicator of success in the profession of medicine, much less any other profession, but clearly there is some room for clarifying the dimensions of success in professions like medicine or law, and while the researcher is not altogether clear, I think some of the research showing the predictive validity of admissions tests in terms of subsequent professional performance as rated by supervisors and others, is very illuminating regarding the fact that aptitude tests of certain sorts do not predict all aspects of professional success.

Dr. **Cooper**. We do not have an aptitude test. It is a medical college admissions test. It is not claimed to be and is not an aptitude test. It measures certain knowledge that the student brings with him and his ability to use that knowledge.

We think that knowledge is related to what is a good physician but this is not an aptitude test, and you and your group may have simple criteria for success in medicine, but let me say that those who have worked in this field and who have tried to assess this have not been able to come up with this.

Dr. **Haney**. May I respond?

My point was not that there are simple criteria of success. It was precisely the contrary, that there are many dimensions of success and that admissions tests do not tap all of those.

Mr. **Weiss**. Mr. Solomon?

Mr. **Solomon**. I think it is fair to say that admissions tests do not often correlate well with future career success.

First of all, they are not intended to. Admissions tests are intended to help predict, along with other criteria, how one fares in the educational program for which one is being considered.

The second problem, as Walt Haney knows as well as I do, is that the problem in research in predicting occupational success is that there is virtually no agreement as to what is meant by occupational success, and the only way which you achieve any kind of research design in order to ascertain that is by very carefully structuring an almost artificial situation.

Is a lawyer more successful if he helps the needy in the ghetto and earns very little money or if he is appointed to the Supreme Court or elected as a Member of Congress?

Who is to judge what success is, especially in terms of all the opportunities open to either a doctor or graduate student or a lawyer after graduate professional school, and most research on predicting occupational success is founded on the criterion problem, not on the predictive problem.

Mr. Weiss. Dr. Cooper, I have a news release that was issued by the organization that you represent as of July 17.

The headline on it is: "New York Law Forces Withdrawal of New MCAT From State."

In the body of that I quote a paragraph:

> Dr. Cooper said the New York law would make it difficult to maintain high quality testing standards, because "there is a limit to the number of relative questions that can be used in the MCAT which has six subsections."
>
> He explained that the test is designed to help predict how well a student will perform in medical school and later as a practicing physician.

Dr. Cooper. This is not an aptitude test. It is a measurement of what knowledge does that student bring with him which is required in modern medicine.

Mr. Weiss. If you are getting into the business of predicting, and it is that kind of statement which appears in the manuals indicating what the test is designed to do but what we are saying is if you are going to make that kind of prediction, then we want to know, the student has the right to know and the public has the right to know on what basis you are making that kind of prediction.

If you don't want to make a prediction, in fact, as Mr. Solomon says, and as I thought you were saying before, that you cannot really make that kind of prediction, OK, then either don't say anything or say that you can't make a prediction, but if you are going to make a prediction like that you will be looked at to substantiate the basis of that prediction.

Dr. Cooper. As we have tried to point out, we do with this very elaborate process which we go through in restricting the kinds of questions we ask. Those questions are now restricted to those aspects of these sciences which are considered to be relevant to medicine and to the practice of medicine.

I went through a long recitation of the very careful way in which that was established for this test over a period of considerable time.

As a matter of fact, the correlations which are in the written testimony showed the agreement among different groups on the value of the tests in establishing those kinds of knowledge and skills which are considered to be essential to physicians.

It will not predict whether this is a compassionate physician. It will not predict whether this physician will take Wednesday and Thursday and Friday afternoons off. It will not predict his clinical judgment.

There are a lot of things about a physician which are very difficult to measure which we are attempting to measure but which you can't make up the qualities of a physician but in which there is no claim here, and really that is a part of the whole predicting process.

Mr. Weiss. What do you mean when you say that the test is designed to help predict how well a student will perform later as a practicing physician?

Dr. Cooper. With regard to the material that that test measures which is knowledge in sciences and the ability to use that knowledge in problem solving which are essential parts of a physician's practice.

Mr. **Weiss**. I must say again that the average reader, regardless of what his level of aptitude or achievement is, reading that statement would draw the conclusion that your test in fact predicts what kind of or how good a doctor he is going to be.

Mr. **Hargadon**. I might try and clarify that.

I do think there is a substantial difference between indicating potential or probability or likelihood of the ability to perform well as a doctor and the phrase which is used here which is quite different, and that is, "Success in the career for which admission is sought."

I guess I see those as two quite different concepts, and that may lead to the difficulty we have here.

Mr. **Weiss**. You may be right, that I think that the utilization of the word success may be vague, but I don't know how that is cured by using predicting performance in later life, later career.

It seems to me that they are both just as subjective, and we may ultimately agree on better phraseology to use in there. It also seems to me that what we are really saying is that whatever the language is that is used in those registration forms or manuals to suggest that those tests are in some way going to be able to forecast performance in the future in a career or in school, that you really ought to be required to put forth a correlation.

Mr. **Hargadon**. I think it would be helpful to us at some point, perhaps in some written form, to note specifically which manuals. You just quoted from a press release, not a manual.

Mr. **Weiss**. We will put it together for you, but this is the quickest thing we could put our hands on to indicate the kinds of things that have caused us concern.

Yes, Mrs. Rice?

Ms. **Rice**. While that particular requirement may be relevant to postbaccalaureate admissions, it has little meaning for undergraduate admissions.

Our tests have but one purpose only—to predict performance in college and only in the first year of college, because our studies have shown, and we have done many, that the prediction of success in the first year seems to be predictive of performance over the entire undergraduate years.

I am not sure, whether you are suggesting that we try to predict success in the career, if the student knows his intended field of work. I am sure you are not, since earlier this morning, you stated that you had no desire to change the content of tests. I think that is what you said this morning to your fellow members of the committee when you were sitting at this table—I have no desire to influence the contents of the tests.

Mr. **Weiss**. Again, I say to you, we may agree on different language to use in there, but whether it is future performance or success in career, if you are talking about in some way gauging how somebody is going to do in school and after in their career, if you make that kind of projection, you ought to have some indication of the correlation between the test and the ultimate outcome.

I am not wedded to any particular language, but I want to be sure that we don't have promises or suggested promises being made which are then not correlated to anything.

Dr. **Haney**. Could I raise a perspective here in that as I testified this morning I have tremendous reservation about many of the specific provisions here, and I am concerned that it might not be terribly fruitful to try to go through specific provision by provision, that it would not be terribly fruitful for the committee or for us here, and that perhaps it would help to turn our discussion to the specific sorts of problems you are trying to get at rather than at the specific language of some of the provisions, and that I know some people have intentions for revising these.

If that sounds like a reasonable suggestion, I would suggest the possibility of inviting the young man who has recently been looking very closely, I know, at testing practices.

Mr. **Weiss**. I appreciate that.

I think that we are going to really proceed with the witnesses whom we have at this point.

Dr. **Haney**. Fine.

Mr. **Weiss**. Mr. Nairn, I think, has been given permission to sit with you to try to respond to questions, but not to give additional testimony to that which has already been given.

Dr. **Haney**. I see.

Mr. **Weiss**. If Mr. Nairn wants to join you to respond to any of the questions that we may have, that is fine.

There has been some question raised and some suggestions made that perhaps the thing to do is to wait until we get the experience of New York. I know that there is division on the basis of the testimony that was given as to the various attitudes toward the New York legislation, even though there may have been no or limited differences of opinion originally.

Given where we are with New York and given where we are with MCAT, for example, I guess Dr. Cooper, the question is really to you, but it really affects everyone else also, because in this instance it's MCAT and it's the dental test people, although we have also seen or heard some possibility that the law school admissions test also may be affected.

That is: Those who administer those tests may decide that they don't want to play with New York either. Now, if in fact what we have is certain broad areas or critical areas within the State being put into a position where the students following that course of study will not be able to take the tests within the State of New York, that really put us in the position where we have no choice except to proceed with national legislation, so that you can't play one State off against another.

Mr. **Solomon**. I just want to say, as far as I know, to the best of my knowledge the law school admission council does not plan to withdraw from New York State.

Now, there may be something that you know that I have not been informed about, but as the agency which does the work I do not have any such knowledge.

Mr. **Weiss**. Somewhere around last Tuesday or Wednesday, there was a newspaper report indicating that the law school people were considering, they had not yet come to a judgment, but they were considering the possibility of following the lead of the medical and dental schools.

Mr. Solomon. Well, sir, we are in daily contact because we have to be with the law school admissions council, and I do not know of that development, and I think we would know about it.

I do know that the law school admission council is exploring all the implications of the New York State bill and trying to figure out as best they can what to do.

Mr. Weiss. Dr. Cooper, do you want to make any comment on that? Doesn't New York find itself in a position where its students in New York who want to study medicine or dentistry are now in a position where they cannot take tests or will not be able to after January of 1980 for admission to those schools.

Dr. Cooper. As I wrote the Governor and Senator LaValle, we have continued to seek ways which would be minimum disturbance to the student's ability to take the test and are looking for some way in which it can be done, but our problem is that the number of questions which we can raise which are already from our test editors being indicated to be new tests.

Questions being developed are just replications of the old questions, and the limits if phrased by that extensive group of people who helped design this test are making it difficult for us if we have to expose every question.

One alternative, of course, I guess, is for us to return to the old MCAT which is much easier to write questions for in large numbers.

We don't feel that it would be the kind of test which would serve the purposes of that new MCAT, and that is the reason we undertook the very extensive kind of effort that we did, to improve the test in our view and in the view of the large group of people that participated in the development of MCAT.

Mr. Weiss. Unless in fact New York State changes its process or you decide to change your mind as of January 1 of 1980, you will not be administering the MCAT in New York State?

Dr. Cooper. I think your two alternatives are the only alternatives, right.

Dr. Haney. I can see how this issue might come up in the short run, but I am very curious to ask, how, given the field that you mentioned this morning, that the fields of knowledge that the new MCAT covers, I believe you mentioned chemistry, physics, and applications in medicine, each of those?

Dr. Cooper. Applications in problem solving.

Dr. Haney. And the applications of those in medically-related fields, leave it at that for now, those are huge areas of human knowledge that, while not infinite, perhaps are so extensive as to make me wonder two things:

One, even if those bodies of knowledge are finite, surely there is enough specific information encompassed in the courses of study in typical colleges in those areas to allow one to develop questions, to appear to; second, though, if those bodies of knowledge are that large, it raises for me a question about the relationship between the finite questions you are speaking of and these very broad knowledge areas that you alluded to as the test measuring.

Dr. Cooper. All I can report on is the views of experts in the field that have had, I think, a great deal of experience in those kinds of restrictions, and I think they are true.

We certainly are not in the position of wanting to thwart legislation or not to serve students. We are in the position of trying to provide a high quality good test, and we do have what I think would be generally recognized as top-notch people in the development of these questions and in advising us on limitations of the questions.

Mr. WEISS. Dr. Cooper?

Dr. COOPER. We can ask questions on quarks in physics. If we ask students about quarks in physics, we certainly are not relating that physics knowledge to what is essential in medical education and in medicine, and we are restricted by the specifications which were drawn up by a very excellent group of people who have wide knowledge and wide competence in this field.

Mr. WEISS. Doctor Cooper, if I recall your testimony, you said that your organization spent about $1½ million on developing these tests, and that it took you the better part of 2 years to do it.

How long had the new MCAT been in effect?

Dr. COOPER. Seven tests have been administered.

Mr. WEISS. And how many tests do you give a year?

Dr. COOPER. We give actually four tests a year for the problems of the Saturday-Sunday thing.

There are really two major dates on which they are given in the spring and in the fall.

Mr. WEISS. And are you saying to me, because I am not clear enough in my understanding of it, are you saying that you use exactly the same tests in each seven of those instances?

Dr. COOPER. No, sir, but we are now coming to the point where we are going to have to recycle questions.

We have just started the new test, and we are building up a test pool but we are coming to the limits of how large that test pool will be and present estimates are we may not be able to go more than four to five more tests before recycling occurs.

If we have to expose every test, then we will never be able to catch up after we have exposed the test.

Mr. WEISS. Again, just so I understand the process in the test, how many questions are on each of the tests?

Dr. COOPER. There are 350.

Mr. WEISS. And can you tell me how many different ones there are on the tests? Are they all different from one to seven?

Dr. COOPER. They are all different.

Mr. WEISS. You are talking about a pool, to begin with, of some 2,300 almost questions?

Dr. COOPER. That is correct so far.

Mr. WEISS. Dr. Haney, have you had occasion to look over or do you have any personal familiarity with the tests that have been given or is there any basis for believing that in fact those tests are in such esoteric areas that there could not be a shifting around?

Dr. HANEY. Shall I respond, and you can respond to my response, if you like.

I must say that I have only read in general of the new MCAT, and I am not familiar with the details.

That said, I would like to raise at least the question of whether or not the limited pool of items that appear to the people who have developed the test in part represent a product of the particular procedures used to develop it, and whether there are any possibilities in the longer run for getting away from the yes or no dichotomy that Dr. Cooper represented earlier?

That is, if there could be other ways of developing equivalent items in a longer term, if certain test development techniques were adopted, but it's very hard for me to ask that question without knowing more detail about the criteria by which the specific items were chosen.

Mr. Weiss. Doctor Cooper?

Dr. Cooper. We had people that did spend a great deal of time looking at this, and there were large numbers of people, both as I tried to point out in the testimony from the experts in testing, as well as those in the subsidy varieties of the title.

It isn't that the questions we use are esoteric, as a matter of fact, quite to the contrary. They are questions which, as I said, we could ask about quarks in physics, but this would be, as far as medical education, would be esoteric.

We will not ask those kinds of questions. The questions are limited by the specifications which try and draw from only those aspects of biology whch are considered by those who are teaching medicine and know something about the related subject.

We are trying to give the student the opportunity to determine or show what he or she knows in the particular areas which are of concern, not about physics generally or about biology generally and, second, only first-year courses.

We have restricted it to that, because we don't want to disadvantage a student who takes a test who, because of his particular interest and because of his curriculum, may not be able to have advance courses by the time he takes the MCAT test.

We don't want to put him at a disadvantage and that is why it's also limited only to the general content of first-year courses.

That was determined with faculty of undergraduate school and premedical advisors from undergraduate schools.

Mr. Weiss. Mr. Solomon, in the course of his testimony, indicated the process that the Educational Testing Service engages in in the development of new questions on a constant ongoing basis.

Are you saying that you really don't do that, Doctor Cooper, that you have contracted with some people on the basis of the best advice and input that you get from people familiar with the field as to the kind and areas that you wanted questions aimed at, and then they came up with a pool of 2,500, 3,500 questions or whatever, and that you really intended for the most part to use that as the basic reservoir, or do you also have the kinds of ongoing efforts to develop new questions?

Dr. Cooper. It's a continuing effort to develop questions. I would also like to point out that I also described a very long process between the time those questions are developed by test writers and their inclusion into MCAT.

That is also, with the number of tests we give a year, that is also a bottleneck. As to these questions, we have them tested in a variety of ways before they ever become a scored part of an MCAT. That is also a problem.

We cannot make a single MCAT so long to include all of the questions. If you have to keep revealing the question, then the student really has to answer two tests, is worn out and there are technical problems about where do you place the questions, at the beginning, the end, are they spread through, and so on.

We take all of these into consideration, but the real issue is in order to develop the final product of an MCAT test, we are restricted not only by the kind of question that can be developed but by the process to assure that that question meets the specifications, doesn't have bias, is unambiguous, and so on, before it will be put into an MCAT.

Mr. WEISS. You are not suggesting that the position of your experts as to why these test questions and tests that have been developed would best be kept in a nonopen system is the uniform opinion of experts in the area of test administration and test creation?

Mr. SOLOMON. This is a very specialized test. It is not like all other tests.

Mr. WEISS. You are not suggesting that experts in the area of test creation and research on tests and test administration are in general agreement that when you come to a specialized area, whether it be in medicine, dentistry, oceanography, or whatever, so unique that in fact you do not give open tests?

I would like Dr. Cooper to respond.

Dr. COOPER. We maintain control over this test. I have not polled all of the test experts in the world with regard to what their views are. Very few of them, I think, as in the case of Mr. Haney really know what the MCAT is all about.

Mr. SOLOMON. The pattern of administration for most of the programs is a little different than that administered for MCAT. In general, I, nevertheless, do want to support Dr. Cooper in his assertion that test development is a laborious process. The problem with the GRE program, I will name one, is that your proposed legislation, sir, would require not simply the development of a few additional forms, but if we were to maintain the service, would require the development of a multitude of forms. As I pointed out in my testimony, we now in the GRE program have special administrations on 116 dates in 9 cities, that is in addition to 5 Saturday administrations and 5 Sunday administrations, in addition in 1978 and 1979 there were makeup tests offered on 38 separate dates, that was 164 different occasions.

Your legislation, sir, would force us to reduce the number of administrations. Right now in that area for the GRE, we make, I believe, five new forms of the GRE aptitude a year. There is a limit, whether you agree or not, to the number of forms that can be developed, and especially in the achievement area where the programs with which we work with committees of scholars and specialists to develop those programs. They are not developed simply in-house by either computers or people who grind out questions.

As far as the programs we are concerned with, even though there are more administrations a year, there is a substantial problem.

Mr. SIMS. Mr. Chairman, I am Albert Sims, vice president for programs of the College Board.

Your reference to Bob Solomon's testimony that there is an ongoing process of item development that is continuous at the Educational Testing Service is not quite true insofar as the College Board's program is concerned. It costs money to develop these tests, these items. We have estimated on the average that it costs us close to $90,000 to develop a new form of the SAT.

In our admissions testing program, we have 15 subject matter tests, which are perhaps more analogous to those used by the medical colleges. We cited the fact—in testimony in Albany, N.Y.—that if we were to expose these examinations each time we offered them, the cost of producing new examinations would be so high that we would probably have to abandon the achievement tests. They are being used less than they were being used for admissions purposes, but there are still some 300,000 students in the Nation who take our achievement tests.

It was in the light of that testimony, I believe, that the achievement tests were taken out of the New York bill. The New York legislature recognized that we could not have afforded to develop those achievement tests and offer them in New York State.

Mr. **Weiss**. Now I understand that you have some familiarity with this question of cost in test preparation. Will you give us your reaction to the argument we have heard today and previously as well.

How much of an additional expense would be required in order to develop all these additional questions if indeed the test became an open test?

Mr. **Nairn**. On the cost issue, there was an important piece of information documented by ETS, studies conducted in 1971-72 and corroborated by later budget information from ETS.

Of the fee paid by a student for a test, although it varies, roughly 5 to 6 percent of the fee paid by the student actually goes to the cost of test development. This cost of test development is to be compared with a proportion in excess of 22 to 27 percent which goes to the profit margin of the test companies. The 22 to 27 percent figure is applied specifically in the case of the College Board, rather of the SAT.

I can cite some specific figures on that. In 1977-78, according to College Board budget, the revenues for the admissions testing program, which includes the SAT, $19,205,998. The expenses for the admissions testing, $14,961,491. This amounts to a surplus of revenues over expenses of $4,244,497, which comes out to about 22 percent. In other years that surplus has ranged up to 27 percent.

Now I will be glad to discuss this with the ETS and College Board officials.

Mr. **Sims**. I would like to have an opportunity to respond to you.

Mr. **Nairn**. This 22 to 27 percent margin it should be noted is only the surplus enjoyed by the College Board on the SAT. The expense figures are those paid to the ETS above expenses incurred by ETS. Per the arrangement, ETS receives compensation for expenses incurred and on top of that they get a fee, which according to ETS internal budget records for 1977 runs in the range of 10 to 12 percent.

So you have the ETS cost and the ETS fee which together represent the expenses paid by the College Board. On top of that there is the 22 percent.

In discussing the cost issue, there has been argument that the profit margin for ETS is lower. They say it runs in the range of 1 percent. The way they come up with these figures, they include expenditures for nontest programs. There is a certain cost for producing the SAT. There is a certain income that generates a surplus which is spent on regional offices, general administrative expenses, et cetera.

After spending this money on other expenses, they may occasionally have a cash surplus left over. This is what they publicly quote as the profit margin for the whole corporation. That may be true for the whole corporation, but the point you have to ask is how much does it cost to produce this test.

Ms. RICE. I hate to interrupt——

Mr. WEISS. Allow me to conduct this hearing. As soon as he is through I will call on you.

Mr. NAIRN. The College Board is 22 percent, and over and above that we have the ETS fee, which runs in the range of 10 percent.

Chairman PERKINS. Mr. Solomon.

Mr. SOLOMON. One hardly knows where to begin.

First of all, Mr. Nairn is correct. He correctly quotes ETS documents when he says there were reports in 1971 or thereabouts that the cost of test development was 6 percent or whatever figure was shown. Indeed, as I tried to say in my testimony, because all of the major programs we serve make very substantial use of existing test forms, because the College Board for example this year budgeted for the development of 24 aptitude and achievement tests, but provided in the year service to students 147 different testing occasions, which means by simple arithmetic they had to have six times the number of forms in stock as the number of new forms developed, indeed one can keep the cost of test development down to 6 to 10 percent of cost.

But if you simply increase the number of forms developed by sixfold or fourfold, you increase the numerator, and the numerator is now 24 percent or 40 percent, and clearly if you do that, one has to increase the denominator in order to meet the increase in the numerator.

Mr. Nairn would argue that is not necessary because he quotes from the College Board budget which shows a difference between revenue and expense. What Mr. Nairn conveniently does is to eliminate all the other services to students including validity services which you, Congressman Weiss, believe there should be plenty of.

You can read our audited statement. Any reader of that will see that indeed there is net income from testing programs, but it is hardly just used for things like administration. One will note in this, there is a very substantial research and development program. In fact, the bulk of our net income from testing programs goes to R. & D., and I might say by the judgment of many researchers it is one of the most outstanding R. & D. programs in the country.

Mr. WEISS. Do you make copies of whatever annual reports——

Mr. **Solomon**. It is published and sent to tens of thousands of people.

Mr. **Weiss**. Would you make it available to members of the committee?

Mr. **Solomon**. Yes.

[Information referred to above appears at end of this day's proceedings.]

Chairman **Perkins**. Mr. Sims.

Mr. **Sims**. On the year ended June 30, 1978—I believe Mr. Nairn's statement was in reference to that—he quotes figures from a summarization of expenses and revenues contained in that statement. Just to give you a perspective on this, in that year the total revenues of the College Board were $52,450,000 and the total expenses were $51,848,000, with a net of $650,000. Now the figures specifically that he quotes, the $19—against the $15 million, are simply artifacts of the way in which the comptroller reports College Board expenses.

I point out to you, he made reference to regional office operations. We have an extensive regional office operation. It provides services to our clients, it answers questions, delivers materials, it holds workshops, provides training for people in the schools and colleges needing to be familiar with the admissions testing program and the changes in it each year. There was over $3 million in expenses in our regional offices. A large amount of that is given over to the testing program. We are an association, and as Mr. Hargadon pointed out to you in his earlier testimony, we are a group of colleges, schools, universities, school systems, and so on. It takes some money to support a cooperative enterprise such as this. It is a fair charge against the various programs of the College Board. I think it is spurious, Mr. Chairman, to quote the kind of figures Mr. Nairn has. The fact is we are a nonprofit organization. We have, over the 79 years of existence of this organization, gotten a total surplus which is hardly adequate in terms of our outside consultants and auditors for the contingencies which this Board must face. It is absurd that we are making profits on this kind of program.

Mr. **Weiss**. Mrs. Rice.

Ms. **Rice**. I am sorry to have interrupted. I wanted to let you know that Mr. Hargadon has to leave, and since he does have to leave, you may wish to address some questions to him——

Mr. **Weiss**. This is an organizational question which I ask so that I can understand more clearly the nature of the association and the relationship. Could you describe for me the relationship of ETS to the College Board? Is it a subsidiary, a total independent agency? Then Mr. Sims spoke about your clients, and am I correct in assuming your clients really are for the most part the same people who are the board, who comprise the board?

Mr. **Hargadon**. I will try to answer that. They are independent institutions, as I indicated. We have a contract, with ETS, a summary of which is included in the materials being passed on to you. We contract with ETS for services. The colleges and schools are interested in certain kinds of tests available to students, and we contract with ETS in the making of those tests. We also have our own panels and examiners who make up the tests, review the tests, but it is a relationship where we contract with ETS to provide many services, and we pay them a fee.

We also reward ETS for efficiency, partly because the aim of the College Board, contrary to whatever mystery must apparently surround it in some people's minds, is interested in providing tests at a relatively inexpensive rate for a large number of students around the country, all of which has led to students passing successfully from one level of education to another.

Mr. **Weiss**. I am sure you are not suggesting the only reason they are passing——

Mr. **Hargadon**. As a matter of fact, I am suggesting that other than the GI bill, the development of national tests of this sort have been the most democratizing thing to take place in this country. I am suggesting that.

What really surprises me about the focus on tests are two things: One is the as yet undocumented and unsubstantiated phrases which are thrown around as to the wide abuse of tests.

I would think the committee would be very interested in showing the public, all of us, the documentation of the so-called wide abuse of tests, either by colleges or anybody else.

Second, all of us in colleges know—and it can be documented—that all of the millions of tests given in the elementary and secondary schools and all the grades given there carry far greater weight on what happens to a student's going on to higher education than do admissions test scores.

Obviously, this legislation does not mean to, open up all the tests and grades to legislative scrutiny or legislation. It is in that sense that the College Board has, through its testing program, has been able to reach out to every region in this country to tap talent among young people regardless of whether or not their parent went to college, regardless of whether they are attending a one-room school or a high school in Beverly Hills. That can easily be documented.

There is no question that the vast number of people going on to higher education in this country is matched by no other country in the world, nor is the diversity of students.

Mr. **Weiss**. Because of the tests?

Mr. **Hargadon**. The tests have played a very valuable role fostering diversity. The tests play a valuable role in that they protect a student against arbitrary actions of a teacher who happens not to like him. If he shows up with a D in chemistry but shows up with an 800 on a chemistry test, you know it may be because he may not have been liked by a particular teacher in that course.

You know there are students with natural ability even though every other measure may show they have not attended a good set of schools previously. The tests made it possible for me before coming here to read the files of a 40-year-old applicant and to be able to enter her to college. It relieves colleges of having to lean on only one variable, her grades, 25 years earlier.

Mr. **Weiss**. You are suggesting in the response that you gave, that is what is under challenge here, are tests. It is not tests. It is the secrecy surrounding the tests that is under challenge, and I do not know why any of the benefits you have described as flowing from the creation and establishment of national tests would at all be reduced if, in fact, you did not have this secrecy surrounding them.

Mr. **Hargadon**. I may draw too many fine distinctions, but I do draw a distinction between the use of the word "secrecy" and "security" for a test.

Sometimes I wonder if people are not trying to have tests so open that in fact everyone will score 100. The fact is, tests are tests, and they do not test much if you have the answers ahead of time. If somebody says here is what I am going to test you on, what kind of test is that? So, there is a sense of security in which you are trying to measure abilities among students. After all we did say in our statement that on the one hand we will try to work to see if it is not possible to meet the ideal, which is to have everybody take the test home after they have gotten their score and so forth. And on the other hand, to promote the validity and security of the test so that it is a fair test and not rigged by somebody ahead of time, by somebody who has gotten copies. It would be nice to have everything, but there are tradeoffs. The College Board will probably work very hard to see if it is not possible—since our main interest is to get students into college, not to keep them out—to see if it is not possible to meet as closely as we can the kinds of concerns you express in your bill, but we also ask you to recognize the kinds of concerns expressed by those who do after all have expertise in this area of testing.

Mr. **Weiss**. I appreciate that. I have two kids who consider I am old fashioned at times, and in listening to your response I was wondering if in fact that does not apply to others than myself. As long ago back as when I was going to law school, open-book examinations were not unusual. There is a school of thought, and I am not suggesting it, that says quite successfully what you are trying to test is thinking and thought processes; that you do not necessarily have to rely on memory but you can provide the tools to implement thought processes.

Mr. **Hargadon**. I cannot agree with you more. I sent out a memo, widely circulated last fall, indicating that it would be great if the secondary schools gave essay examinations while students were still at that level so they would not hesitate to write when they get to college. I am not a great multiple-choice person. We all want the kind of qualities that you have suggested. At least we would like to be able to perceive and recognize them, but there is a very practical matter. I give open-book exams at my classes at Stanford, but the faculty and students operate under an honor code system. We do not have a national honor code system for taking tests.

Mr. **Weiss**. Again, giving what has to be an open secret, to wit the fact that there are tutoring schools which specialize in getting hold of the questions, then preparing their kids with the answers, I just wonder why you think those tests are so secure, why in fact you do not agree they are secure for a lot of people, but for a lot of people there is a built-in advantage because of the system.

Mr. **Hargadon**. Maybe you know, I do not know, and I will simply guess that of the millions of testtakers in this country every year, particularly the SAT, only a small fraction have gone to a coaching school. The argument could be if, in fact, a few can breach security, then there ought not to be any security at all. I do not accept that argument. I do ask you to consider, in all the publicity about the coaching schools, the possibility of what some people term a coaching school may turn out to be 10 weeks, 4 hours a week, of teaching. That does not say anything about whether abilities are developed in those schools or not. It may say something about what is not being done in the secondary schools turns out to be teaching for money over a long time in the coaching schools. There is quite a difference between this and what they tend to imply in their ads; namely, there is a technique and we—the coaching schools—can coach you on the technique.

Mr. **Weiss**. I don't know if you have any studies which support that conclusion but, regardless of which way it really is, it's the kind of argument in the abstract that I can't win, because I certainly cannot prove to you that in fact that may not be the case, but it certainly disproves the proposition that in fact attending these schools in no way influences which is what some of the suggestions that have been made, that it's just not possible for people to be coached for these examinations.

Mr. **Hargadon**. I am going to have to excuse myself. I am going to try to arrange an operation for my 16-year-old son with a doctor whom I hope knows a lot about medical science and very little about quarks.

Mr. **Weiss**. Thank you very much for your participation.

I appreciate your patience with us.

Mr. **Nairn**. I would like to comment on the response of Mr. Solomon and Mr. Sims' argument about costs, and I would like to begin by bringing in an opinion of a third party, namely, Forbes business magazine, which I am sure the members of the committee are familiar with.

In 1976 they did an article on the finances of ETS, entitled, "Pleasure of Nonprofitability."

According to Forbes, nonprofit in theory, Princeton's Educational Testing Service, in fact, shapes up as a tough, aggressive and even dynamic growth business.

> If Princeton, New Jersey's Educational Testing Service were a public company and not a self-contained tax-exempt nonprofit organization, it would probably have long since emerged as one of the darlings of Wall Street.
>
> The knowledge industry has generally manifested more promise than performance, but ETS has demonstrated all the performance any promoter could wish. Over the past 30 years it has easily racked up a record as one of the hottest little growth companies in U.S. business.

ETS has not, through profit growth, served to generate a sizable pool of retained earnings.

The article continues with extensive charts and documentations of the growth in retained earnings, profit of ETS. I would suggest that ETS has been able to acquire this financial reputation by a system of charges for its test which is somewhat different than that described by Mr. Solomon and Mr. Sims.

In fact, as Forbes points out, very large profit margins are obtained in the specific testing programs that the ETS administers.

As I discussed before, there is a 22-percent profit margin.

They go on to discuss the way they spend these profits, and they argue that the way they spend those profits are worthwhile. That may or may not be the case, but the point is ETS and the College Board are given their nonprofit charters for the specific purpose of administering testing programs.

That is the primary purpose for their existence as a corporation.

The question that pertains to this legislation is how much does it cost to administer the tests, and how much income are they getting?

They are generating substantial profits in the administration of each of these tests.

Mr. Solomon went on to suggest that under conditions of a truth in testing bill, the test development costs would multiply so much that they would be far out of proportion to what they are now.

For the opening legislation, ETS submitted to Senator LaValle specific estimates of how much the bill would increase test development costs.

In the case of the SAT, for example, ETS estimated by its own methods that the bill would increase SAT development costs by $1.92 million.

This amounts to less than a third of the surplus of profit margin on the SAT program.

Mr. Weiss. Nationally or just in New York State?

Mr. Nairn. This was in order to comply with the provisions of the New York bill which is the same as the provisions of your bill, as far as disclosure questions, so the estimate would apply in both cases, so this is EST own estimate.

We could argue a number of ways that estimate is inflated, but according to their own, it would amount to $1 million.

You would have to obtain and look at those kinds of figures across the board for all of ETS other test programs in order to fully document this issue.

That kind of documentation is not included in the ETS annual report which Mr. Solomon offered to submit.

If the committee is interested in the financial issue, they should also request ETS project operating statement where they for their own internal accounting purposes list specific expenses as well as profit margins on each of their programs.

With that kind of information on hand as well as information on revenues from each specific testing program you could then look at the profit margins on each program which would be very helpful for the purposes of this legislation, so that might be requested in addition to the annual report.

Mr. Weiss. Mr. Solomon?

Mr. Solomon. We, as I say, issued the annual report. We will be glad to leave this copy with you.

This includes a statement of assets and liabilities since the inception of ETS.

The wealth that Mr. Nairn is talking about amounts to $4½ million in liquid assets, and about $24 million in plant and equipment and land.

ETS started with a very small plant. I don't know the exact amount. Mr. Nairn probably does.

To start ETS off, there was no endowment. We managed on the basis of the income from test programs.

Most of that net that Mr. Nairn refers to has gone into buildings, as I say, $24½ million.

Our total annual operating expense is over $90 million, of which we have a $4½ million liquid reserve which, by any business standard, that is extremely low.

Furthermore, in this same annual report, you will see a 5-year summary of revenue which shows the income from College Board, from other graduate programs, which shows research and development services and, in other words, gives us the very breakdown that we are talking about.

Mr. Nairn financially speaks of the estimates we gave with regard to the SAT development, and he quotes the figure correctly.

The point I have tried to make is that that is one test, and in fact for some other testing programs that ETS administers, the cost of limitations of your proposed legislation, sir, are even more horrendous than they are for the college boards.

I cited the problem with the GRE program with all its special administrations more than those for the College Board, so it isn't just $1.2 million. It's $1.2 million for SAT.

It's the achivement test for the College Board which is a much more serious problem. The GRE advanced test, the aptitude tests, the Miller analogies test, medical college admissions tests, dental exam and each of those has a price tag next to it which is substantial, and your legislation, sir, would inevitably, even if everyone could find a way to do it and I am not sure they could, would, either, in terms of trade-offs, either increase costs to students or reduce service to students or affect the quality of the program.

I would submit to you, sir, although I think your motives are of the highest, I would submit to you that some of the people supporting this legislation indeed do want to bankrupt the tests and get rid of the tests.

Mr. WEISS. Let me just suggest or invite you, whatever information you have, whatever projections you can make as to financial or expense costs, increases in the cost of preparation, the cost of giving, whatever on any angle of the tests, I would welcome that and the committee would welcome it.

I urge you to please submit that to us.

I would also urge you to submit to us the kind of information that Mr. Nairn was talking about which is the project development information to indicate what in fact goes into coming up with the conclusions as to what the increased costs would be, because perhaps that way we can look objectively at least to see where the dollars and cents lie both as far as income and outgo are concerned.

[See materials at the end of this day's proceedings.]

Mr. SOLOMON. I understand what you are saying, but I think there is one other point that needs to be made.

Mr. Nairn argues that what agencies such as the College Board ought to do is just develop tests and administer them and collect fees and do nothing more.

I submit to you that if the College Board did that, we would now be having hearings held by you to know why the College Board is not giving advice to colleges on the appropriate use of tests and conducting validity studies on tests, why the College Board isn't providing publications to students and better understand the test, such as the free SAT booklet published for taking SAT's.

In other words, you, among others, would probably be among the very first who would be genuinely concerned with the fact that students do have to have information which is, of course, the principle behind your proposed legislation.

Now, we have a situation where in fact that is what the College Board is doing, and Mr. Nairn is saying that is a bad thing to do. In fact, they shouldn't do any of those things. They should just administer tests and score them and nothing else.

Mr. **Weiss**. I really don't want this to be a focus on what Mr. Nairn wants.

It is what I want and what the committee wants, and what we are trying to do and the purpose of the hearing is to try to establish a factual base for coming to some conclusions as to what we ought to do.

One of the arguments that we have heard repeatedly, both at this hearing and previously, is that if my legislation is adopted the costs are going to be just astronomical, and it will be impossible to do it.

Well, OK; maybe they will and maybe they won't, but at the very least, never mind Mr. Nairn, what I would like is for you to substantiate those statements.

Show me where in fact the costs are going to be and show me what it costs you now to develop the test on an individual test basis, and what the difference will be when this legislation goes into effect, and then maybe we can in fact come to some understanding as to why in fact my suggestions are so outrageous or why they are going to be so expensive.

That is the basis for doing it.

One specific question——

Dr. **Cooper**. May I say, I also am afraid I have to go.

I have to travel and I apologize. We have been here a long time.

Mr. **Weiss**. You have, indeed.

I appreciate the fact that you have been here for the bulk of the day, and I thank you for your patience and your cooperation, and I urge you not to put our students in New York into the position where they will have to go out of State to take examinations because I tell you as a word of advice, that that probably will create more of an impetus for more legislation than anything else that you could think of doing.

Mr. Solomon, you had said in the last round of responses that most of ETS's so-called profit margin, the surplus, went into research and development.

Is that research and development for test purposes or followup validation or is it general research and development?

Mr. **Solomon**. It's both. Some of it is directly connected with tests, test development, new kinds of test development.

I said in my testimony, I made reference to the fact that we are doing research development on tests which measure personal accomplishments, other characteristics that are not measured by the current ability tests.

Some of the research money goes into that, but in addition to that, if one looks at ETS's charter, part of our charter obligation is to conduct research to improve measurement generally, not just the programs that we happen to administer, and so indeed some of our research is not related to the testing programs, but is related more broadly to questions of educational development and growth and better understanding that process and applying that knowledge to solutions of other educational problems and, indeed, part of our research and development goes to that and it's well known that it does.

Mr. **Weiss**. Do you have a breakdown as to what percentage of it goes for general research, educational research, development and for testing?

Mr. **Solomon**. I don't think we have ever tried to break it down that way, and I would not be sure we could, because that suggests that there is some sort of sharp division between one kind of research and development and another, and it is what one learns in basic research, one then applies in more applied situations.

Mr. **Weiss**. You see the problem that it presents to some of us in getting into the responses to the questions as to how much more it would cost the individual student to take the examination, or how fewer oportunities a student will have to take the examinations, is that we don't know how much of the money really is being utilized for nonpriority purposes, if in fact your prime purpose is for it to provide for tests.

Mr. **Solomon**. Again, Mr. Weiss, our charter is not that narrowly written and, therefore, the question you raise is: Who is to decide what the priorities are?

Now, we make no secret of our work. It's well known nationally in the research community and even in the educational community; our research reports are contrary to some things that have been alleged are published.

We distributed over 10,000 research reports this last year, but I come back to the central point that our charter obligation was to provide testing programs and services, conduct research in the area of measurement and provide advisory services and field services in the broad area of educational research.

We were not created originally to be narrowly a test building and test selling organization, but rather were created, and I had nothing to do with the creation, but the inspiration of those that did was to create an agency that would serve the educational community in that broad a way.

Mr. **Weiss**. You are saying that there is no way that you can break down and give us a figure as to, out of those three specific responsibilties, how much of your money goes to testing?

Mr. **Solomon**. That breakdown would be artificial and misleading but exists.

Mr. **Weiss**. But exists?

Mr. **Solomon**. No; no, sir. I didn't mean to imply that it did.

Mr. **Sims**. I want to correct a statement that had been made about the charter of the College Board.

We are chartered by the Board of Regents of the State of New York to act as a forum for the discussion and exchange among educational institutions and secondary schools, of educational standards, an important associational function, and to provide certain kinds of educational services.

The people who wrote that charter, in their wisdom, never in any way restricted our role, even primarily, to testing.

Let me say something else that is more on my mind than that miscellaneous piece of information.

I have testified now twice in Albany, once in Maryland, and now here with respect to this kind of legislation.

The point just made by Mr. Nairn is repeatedly made with respect to the Educational Testing Service.

In fact, in Maryland the bill was labeled a bill to regulate the Educational Testing Service.

I suggest, sir, that that is a misconstruction of what the problem is, if there is any.

The College Board, insofar as its programs are concerned—and we have a wide variety of tests that we sponsor and a lot of that activity we conduct outside of ETS.

ETS is in full control of its programs. but not ours. It decides what programs there will be, and appoints the committees of teachers and professors who will formulate the tests, who incidentally see that those tests are relevant to what is being tested.

We, the College Board, set the fees. We are the ones who monitor the quality of service for our programs at the Educational Testing Service.

We are the ones concerned, through our trustees and our finance committee, with the reasonableness of costs for the services performed for us at the Educational Testing Service.

We are the ones who will decide whether the SAT will be offered five times or seven times or two times and in what circumstances and where, because that is a policy affecting one of our programs. Unfortunately, few understand that the Educational Testing Service is a separate corporation and they can take care of themselves; as Bob Solomon has given some indication here today, they are able to do just that.

The issue is whether testing policy should be substantially changed in ways that your bill suggests and that is a policy consideration and the final answers as to whether it can be done and how it will be done will be rendered by the College Board's Trustees of the College Board Association.

Of course, we, the College Board, will be affected by the testing legislation, but it seems to me that what has been happening here is that a strawman has been set up to attack, namely, the Educational Testing Service when that is not really the issue of proposals before you, as I understand them, at least.

Mr. WEISS. I am curious about one of the comments you just made and, again, misapprehension on my part, and you correct it for me.

Are you saying the development of questions for the tests and the hiring of professionals in the field to do that, is done not by the Educational Testing Service but done by you?

Mr. SIMS. No; I am saying that the appointment of committees and the payment of committees—and we have test development committees in for every College Board test including the SAT and the achievement test—is made by the College Board. We select the faculty members or school teaching faculty that are on those committees, because we want to get tests that are relevant and fair for the students in the country.

Those test committees work with the ETS test development staff. Bob Solomon reported in his earlier career he was one of those test developers. We work with them, and our staff also sits in on those meetings from time to time to monitor the kind of activity that is going on, so in a real sense we take a policy and substantial responsibility for the nature of the tests.

Mr. **Weiss**. To answer my question specifically, the tests are developed by the Educational Testing Service, are they not?

Mr. **Sims**. That is right. The standards with respect to them are set by our committees, and our committees pass on the standards for developing the questions.

Mr. **Weiss**. Again, so I understand this relationship, for background I read somewhere along the line that the Educational Testing Service was in fact created under the auspices of the College Entrance Examination Board.

True or not true?

Mr. **Sims**. Not quite true.

Mr. **Weiss**. Tell me what is true.

Mr. **Sims**. The College Board was one of three sponsoring organizations in 1947, 1948.

Mr. **Weiss**. That matters?

Who were the others?

Mr. **Sims**. The American Council of Education and the Carnegie Foundation for the Advancement of Teaching.

Mr. **Weiss**. What percentage of the work of the Educational Testing Service was done for the College Entrance Examination Board?

Mr. **Solomon**. About 45 percent.

Mr. **Weiss**. From the beginning or over a period of time?

Mr. **Solomon**. It has varied, gone up or down.

The estimate I gave you is about what it is now.

Mr. **Weiss**. Because, again, Dr. Sims, it seems to me that you would understand that, I would hope that you would, it's not a matter of attacking a man, that regardless of how you set ultimate policy, if in fact we are told because we want insurance that the secrecy or security aspects of the tests are removed to the extent that people can know what examinations they take, and so on.

If that is done, it will cost a great deal of additional money, because ETS says that they are going to have to charge more for it.

Then it seems to me that that is a matter of scapegoating. That is a matter really of finding out what the facts are, the relevant facts. Is that not so?

Mr. **Sims**. Yes; it is so, but the point I was making was that the decision as to how you would comply with such legislation which will undoubtedly be some combination of increased costs and reduced test availability is a decision of the kind that would be made by the College Board itself.

Dr. **Haney**. I would in a way like to take Mr. Sims' point even a bit further than he took it in that it seems to me the issues and concerns behind the proposed legislation we started out discussing here today go far beyond ETS and certainly far beyond College Board.

In addition, they raise important educational questions about the relationships between those commonly used tests, secure tests, or secret, whatever word you choose, and what goes on in our schools and colleges.

I hope that the committee, if it wants to pursue these issues further, will call on not just the people who are primarily responsible for these tests but educators much more broadly, because my ultimate concern really is at root, it is clearly concerned with College Boards and the others, is what this has to do with learning and teaching in our educational institutions.

If I may make that suggestion, and in closing perhaps a little historical perspective in that the debate today, much of this got caught up in the costs of what ETS entails and what they pass on to College Boards, et cetera, and I could envision this hearing concerning regulations, of General Motors.

We are getting into cost issues, accounting issues, when in fact the real substance concerned is education.

There is a real tension between current procedures and the current means by which test validity is determined and on security as it is now construed, and there seems to be a real tension there.

The historical antidote I would like to bring to your attention is that the times change. What we can do in education does change.

I hope we can have a broader view. I would note simply prior to 1958 it was deemed inappropriate even to tell the individual students their scores on the SAT. There was much controversy about whether it would be possible or desirable to give students even knowledge of their own scores on the SAT's.

There was much doubt about that at the time. Procedures were worked out so that that was accomplished, apparently not without too many unforeseen or detrimental consequences on individual students, even though that openness in testing provisions was hotly debated at the time.

Mr. **Weiss**. Ms. Rice?

Ms. **Rice**. I would like to make a couple of comments.

I am delighted that we have gotten back on the issue of education and away from some of the other issues.

I would very much like to call your attention to the top of page 12 of Mr. Hargadon's testimony, and ask you to compare the costs of our services over a 20-year period. Those costs have gone down in real terms over that time. We now charge in 1978–79, $1 more than we charged for our tests 20 years ago.

Mr. **Weiss**. How many students took the test in 1958?

Ms. **Rice**. I'm not sure. Admittedly, costs are related to volume; for example, the volume went up in 1968.

The midpoint between 1958 and 1978, and our costs went down to $4.50.

My main point, Mr. Weiss, is that I have spent my entire career trying to improve education in this country, and particularly the lot of those people who have been and continue to be the "have nots" in this society.

I would not be working for this organization if I didn't feel and know that some of that so-called profit that has been discussed was not being devoted and dedicated to help the helpless.

I am not sure that you would like us to find no difference between the income and costs of our tests. If that were the case, then we could not give away nearly $300,000 in fee waivers annually to poor kids.

We could not contribute nearly $100,000 a year to an educational opportunity center here in the Anacostia part of Washington.

We could not train the people in the colleges and universities to be sensitive to the needs of the disadvantaged and to change their old admissions policies to move away from objective criteria, and to consider what other strengths the poor and the forgotten and the racial minorities have been able to bring to higher education.

Ms. RICE. I guess I am a litle suspect—I am extremely suspect of the legislation for the policy reasons we have outlined in our formal testimony. But also I am also concerned that the consumer groups supporting this legislation have not been that interested in the poor, because if you are poor, you seldom are a consumer, Mr. Weiss, I think a lot of the consumer legislation that we have has been directed to middle-class Americans.

I simply want to say this organization has been devoted to some of the great ideals of this committee and ideals of yours.

Moreover, we tried to focus our testimony on educational issues not testimony on the fiscal costs. The real costs are those we would have to pay in terms of this legislation in as it redirects the Federal role in education.

There are some constitutional questions and protections that this legislation seems to defy. And there are still other problems that we tried to stress in our testimony.

Mr. WEISS. I am aware of the good works the College Board is engaged in, but just yesterday I sat as a member on the Subcommittee on Employment Opportunities, and we had before us the Chair of the Equal Opportunities Commission, Mrs. Norton.

In response to a question I raised about sex discrimination, affirmative action within the universities, she told me that the colleges and universities were the great scandal of this Nation; that private corporations and presidents of private corporations would not dare to react as cavalierly and callously to questions about affirmative action and failure to meet affirmative action requirements.

So I must tell you that basically it is the same people in both instances. I am not overwhelmed with the commitment the colleges and universities have as to equal opportunities—maybe sometimes yes, but not everybody everywhere.

Ms. RICE. As I sit here at this table as an officer of The College Board with all these gentlemen, it should be clear that I am the only person who is both a woman and a black. We are the only major educational association in Washington which has taken such a step—I cannot speak for all colleges and universities—but I can speak for the Board and our record has been admirable compared to most.

Mr. WEISS. Nor do I understand the kind of conspiracy suggestions implicit in a memo you gave to Ms. Chisholm in which you close it out:

> Ultimately, the Weiss bill is a mysterious piece of proposed legislation. There are no self-evident reasons for its introduction, no known scandals to correct, and no

apparent basis for the Federal intervention proposed. All of which suggests there may be a hidden agenda that will in time be revealed.

I do not know what the hell that means; and I wish you would tell me.

Ms. **Rice**. Much of that memorandum was addressed to a provision which you have deleted from your legislation. You had a section in the draft legislation which, in my view, would have set back efforts, and I think in Mrs. Chisholm's view also, to diversify institutions of higher education.

You asked for the equation or formula used by colleges and universities to admit students. I was not impugning your intentions. I have great respect for you, but that provision touched a deep chord in me when it was in the legislation. We have just gone through a long effort on the *Bakke* case and I was a little bit concerned that this specific provision, intended or nonintended—and I assume it was unintended because you dropped it—would have placed far too much reliance on objective criteria. That concerned me. The Supreme Court I think reaffirmed in the *Bakke* case the understanding that we have had in many other cases, that institutions can choose who to teach, how to teach, and who shall teach. I just believe, sir, that memorandum was written because of those concerns. I am delighted that you removed the provision.

Mr. **Weiss**. The memo goes to the whole legislation. What you are speaking of, we dropped because it was secondary. I appreciate your suggestions along those lines, but I just think that it is not necessary or should not be necessary in the course of this kind of discussion to look for hidden agendas. I have none; I do not think you have or they have. We are dealing with a very serious and profound question and we are trying to deal with it as to what we think is best for this Nation.

At this point perhaps the thing to do, because it is getting late, is to ask if anyone has any final brief comments to make so we do not leave anyone not being able to respond.

Mr. Nairn.

Mr. **Nairn**. First you asked about the College Board-ETS relationship. The president of ETS is ex officio president of the College Board; the president of the College Board is an ex officio member of ETS.

Second, there has been some discussion of the services the College Board provides to its member colleges as being a worthwhile activity, and I am sure it is. The College Board has two sources of income: One from membership fees from colleges; the other test fees. If they want to provide more extensive services to colleges, that is commendable.

Mr. Solomon briefly stated all ETS research studies are published. There are also certain categories of ETS reports not revealed, such as the series of reports known as test analysis.

Finally, there has been some reference to the way that the profit margin is spent to provide free waivers and so on. It would be helpful if ETS and the College Board provide specific breakdowns to show revenues. Also, how much of it went to lobbying the truth-in-testing bill in New York; how much goes to support the Henry Chauncey Conference Center, which recently opened its new swimming pool; and how much goes to other expenses to support the high lifestyle of ETS.

I am proposing the profits all be laid out in the open for inspection by any interested member of the public. I am sure ETS will cooperate by providing the project operating statement and other relevant documents.

Dr. Haney. I am not nearly so interested in ETS, the College Board, and their interrelationship or what have you, as the larger question, which I really hope this committee will address more thoroughly, namely the relationship between testing and assessment, particularly with respect to secured tests and the educational process in this country. It seems to me there are two fundamental concepts involved, educational and matters of fairness and justice. It seems to me those two perspectives have been notably absent today, and I would urge the committee in its future deliberations try to involve others to talk and examine more thoroughly the consequences of secret testing, and I would hope since I truly do believe there is an issue of fairness of justice involved in requiring as matters of public policy, students to take tests, the results of which they do not see in any detailed way, I think there is an essential fairness issue involved there. That this committee will try at least to some extent to draw on at least some representation of the students and children who might be affected or at least their representatives.

Mr. Weiss. Mr. Solomon.

Mr. Solomon. I want to thank you for your patience in listening to us.

Mr. Weiss. Mr. Sims.

Mr. Sims. I think I want to agree with Dr. Haney. The essential issue is a policy and social issue. It involves a question of whether in the broad sense why is it necessary for government to intervene in regulations in this field, and it also seems to me it is a question of social and educational values. There is implicit in the agenda for action certain values that need to be made explicit and discussed, and they need to be deemed important in terms of serving the public at large. It seems that is as yet a very controversial issue.

Finally let me say the College Board has faculty who oversee our program and discuss the relevance of testing to education. We have students sitting on those committees, we have an apparatus for action which is proved and can accommodate itself to change very fluently.

Mr. Weiss. Ms. Rice.

Ms. Rice. Thank you and other members of the committee for giving us this opportunity.

Mr. Weiss. Thank you for your participation.

I have one request, Mr. Goodling had indicated he would like to submit written questions, and I am sure other members of the committee would like to do that. If you will respond to those. Thank you very much.

Committee stands recessed until tomorrow morning.

[Whereupon, at 3:10 p.m., the subcommittee was adjourned, to reconvene Wednesday, August 1, 1979.]

[Additional material submitted for the record follows:]

association of american medical colleges

JOHN A. D. COOPER, M.D., PH.D.
PRESIDENT

202 828-0460

August 10, 1979

Honorable Carl D. Perkins
Chairman
Committee on Education and Labor
U. S. House of Representatives
Washington, D. C. 20515

Dear Chairman Perkins:

On Tuesday, July 31, 1979, I testified before the Subcommittee on Elementary, Secondary, and Vocational Education about two bills on standardized testing, H.R. 3564 introduced by Mr. Gibbons, and H.R. 4949 introduced by Mr. Weiss et al. Enclosed is a copy of my testimony.

While H.R. 3564 poses some serious problems for medical education, H.R. 4949 threatens the very viability of the New Medical College Admission Test (New MCAT). This examination provides information critical to the fair and impartial selection of students to study medicine.

The medical schools of this country, the sponsors of the New MCAT, are concerned that these bills, and especially H.R. 4949, will not receive proper scrutiny and that their destructive consequences for medical education will not be recognized. The danger is particularly acute since they are being promoted under the deceptive banner, "Truth in Testing."

I, therefore, earnestly request your careful attention to these considerations:

- No documentation of the alleged problems has been presented. Contrary evidence exists.
- MCAT test scores are not the sole criterion of acceptance. Formal studies show them to share about equal weight with grades, interview ratings, and letters of recommendation.
- The provisions of the bill will not achieve the so-called reforms. In fact, they are likely to be counter-productive.
- The provisions are highly intrusive into education and the admissions process at a time when the Congress is particularly sensitive to such legislation.

Suite 200/One Dupont Circle, N.W./Washington, D.C. 20036/(202) 828-0400

- Action at the federal level prior to an evaluation of the consequences of a similar New York law is premature.
- The AAMC just spent $1.5 million in revising its admissions testing program to increase its relevance and utility. The New MCAT incorporates all the objectives of these bills regarding public awareness and protection.
- The highly specialized nature of the New MCAT and the restriction on questions to assure relevance to medicine is not compatible with continuous disclosure of test material.
- The New MCAT Program is completely accountable to the educational community it serves and its policies and procedures are regularly reviewed.

The members and staff of the Association of American Medical Colleges will be very pleased to discuss any of these matters in greater detail with you or your staff. Thank you for your serious attention to these concerns.

Sincerely,

John A. D. Cooper

John A. D. Cooper, M.D.

MAJORITY MEMBERS:
CARL D. PERKINS, KY., CHAIRMAN
WILLIAM D. FORD, MICH.
IKE ANDREWS, N.C.
GEORGE MILLER, CALIF.
AUSTIN J. MURPHY, PA.
BALTASAR CORRADA, P.R.
DALE E. KILDEE, MICH.
PAT WILLIAMS, MONT.
AUGUSTUS F. HAWKINS, CALIF.
MICHAEL O. MYERS, PA.
RAY KOGOVSEK, COLO.

(202) 225-[illegible]

CONGRESS OF THE UNITED STATES
HOUSE OF REPRESENTATIVES
COMMITTEE ON EDUCATION AND LABOR
SUBCOMMITTEE ON ELEMENTARY, SECONDARY,
AND VOCATIONAL EDUCATION
B-346C RAYBURN HOUSE OFFICE BUILDING
WASHINGTON, D.C. 20515

MINORITY MEMBERS:
WILLIAM F. GOODLING, PA.
JOHN H. BUCHANAN, JR., ALA.
ARLEN ERDAHL, MINN.
DANIEL B. CRANE, ILL.
JON HINSON, MISS.
JOHN M. ASHBROOK, OHIO,
EX OFFICIO

November 9, 1979

Dr. Robert J. Solomon
Executive Vice President
Educational Testing Service
Princeton, New Jersey 08541

Dear Dr. Solomon:

On July 31st you testified before the Subcommittee on Elementary, Secondary, and Vocational Education on H.R. 3564 and H.R. 4949. During the course of that hearing you were asked to submit for the record several pieces of information. After reviewing the transcript and the materials submitted by ETS, we feel that you did not respond in full or, in some cases, at all to some of these requests.

We would therefore appreciate your responding again to the following requests for information:

1. A copy of the ETS Project Operating Statement;

2. A complete statement of the costs, revenues, and surplus for each ETS test covered by H.R. 4949, including an account of test development cost and how the surplus is allocated;

3. The additional cost, if any, on a test-by-test basis that would be incurred if H.R. 4949 were enacted including a specific breakdown of how and why such additional costs would occur; and

4. A full description of ETS' research and development budget including a statement by percentage of the amount of research and development costs which have direct application to current testing programs.

We would appreciate receiving a response to these points in writing by November 22nd for inclusion in the hearing record.

Thank you.

Sincerely,

Carl D. Perkins
Chairman

EDUCATIONAL TESTING SERVICE PRINCETON, N. J.

Robert J. Solomon
Executive Vice President

November 14, 1979

The Honorable Carl D. Perkins
Chairman
House of Representatives
Committee on Education and Labor
Subcommittee on Elementary, Secondary,
  and Vocational Education
B-346C Rayburn House Office Building
Washington, D.C. 20515

Dear Chairman Perkins:

Thank you for your letter of November 9 in which you request certain information from ETS on behalf of your Subcommittee.

During the past few weeks, Mrs. Alice J. Irby, our Vice President in Washington, has been in conversation with members of the Subcommittee staff on several occasions. It is my understanding that they agreed upon a range of materials to be submitted as a follow-up to my testimony. Mrs. Irby has informed me today that the materials agreed upon in these conversations have been delivered to the Subcommittee (attention of Mr. Jennings and Ms. Kober) under a cover letter dated November 13, 1979. I trust these materials, although they are not precisely the same as requested in your letter, will indeed serve the purposes of the Subcommittee.

May I take this opportunity to thank you again for your careful and thoughtful consideration of our expressed concerns about the proposed legislation.

Sincerely,

Robert J. Solomon

Robert J. Solomon

EDUCATIONAL TESTING SERVICE WASHINGTON, D.C. OFFICE

Suite 300
1800 Massachusetts Ave., N.W.
Washington, DC 20036
•
202-659-0616

ALICE J. IRBY
VICE PRESIDENT

November 13, 1979

House Subcommittee on Elementary, Secondary,
and Vocational Education
Room B346 C
Rayburn House Office Building
Washington, D.C. 20515

Attention: Mr. Jack Jennings, Ms. Nancy Kober

Gentlemen:

In response to your request for information to augment Mr. Robert Solomon's testimony before the Subcommittee on July 31, 1979, I am enclosing statements which, together with the previously submitted document entitled "Educational Testing Service, Finances..Where the Money Comes From and Where It Goes" provide data on sources of ETS income and on test development costs.

The first table gives a picture of ETS revenues at ten-year intervals from 1959 to 1979. From it, one can ascertain the growth in categories of ETS programs and the changes over time in the composition of total revenues by program area. Two points are of interest:

1. College Board programs, while the largest single component, have declined proportionately since 1959. The $45.5 million in 1979 includes all College Board programs of which the Preliminary Scholastic Aptitude Test and the Admissions Testing Program (SAT and Achievement Tests) are only two.

2. Professional and Occupational Programs have doubled as a proportion of the total, and Graduate and Professional Programs grew from 12% to 17% of the total.

It has been suggested that ETS should provide revenue minus expense and fee on each of the programs it serves which are affected by the legislation. Doing so would not only provide an incomplete financial picture of each program but would also be misleading as to full program costs.

Most of the programs affected by the legislation are conducted under contract with independent sponsoring groups. These sponsors have expenses which must be met in addition to ETS's charges. Data reflecting total revenues and costs are the property of these independent clients. Often, ETS does not have access to these complete data. Even when it does, it does not have the authority to decide to whom it will be reported. The sponsors make regular financial reports to their memberships and to various governmental agencies.

ETS publishes and distributes routinely financial reports concerning its own financial affairs which are prepared in accordance with generally accepted accounting standards and certified by independent public accountants. A copy of the 1978 Annual Report is attached.

ETS' reluctance to release more detailed cost information stems also from the fact that it submits proposals in competition with many other organizations, both non-profit and profit, to offer testing, research and related services. Contractual arrangements with each client are determined through negotiation and depend upon the specific circumstances of each situation. Factors which must be considered include the nature of the service to be provided, the transferability of the methods used in designing services for the same or other clients, the opportunities for research or exploration of new methodology, the length of the contract and the financial stability of the client. Sometimes, ETS enters into contracts with a very low fee or none at all to encourage development of a service which meets a pressing social or educational need. Making public the financial details of each individual activity would give competitors considerable and unfair advantage in bidding for contracts and substantially restrict ETS' ability to negotiate contracts appropriate to particular circumstances. To our knowledge, the Congress does not generally require corporations, non-profit or profit, to release specific data from which cost rates and confidential client data can be derived. To do so in this case would be contrary to customary practice.

B. The second table is reproduced from Attachment 6 of Mr. Solomon's testimony. It gives examples of projected increases in test development costs, given certain assumptions about disclosure of tests. From the table, one can estimate increases in costs and consequent likely increases in student fees, given different assumptions about the number of new forms produced annually and the number of test takers.

1. The current average cost per test form in GMAT is approximately $92,000. Actual costs vary a bit depending upon the scheduling of work. GMAT introduces two forms per year. Thus, the production of six new forms

(assuming eight national administrations are continued) would add $552,000 annually to a current base of $184,000 or an increase of 300%. If the number of test takers remains constant, the increase in fees would be about $2.00 from test development cost increases alone. This, of course, does not include other administrative costs, such as printing more tests in smaller quantities per test, mailing of forms to students, copying of answer sheets, all of which might lead to increases in fees.

2. The same calculations can be made for the six GRE advanced tests affected by H.R. 4949. Currently test development costs for these exams approximate $300,000 per year. Assuming they are administered ten times in national administrations, the test development cost increase would approximate $2,700,000 or 900%. If the number of test takers remains 70,000, fee increases of $38-$39 could result.

These data assume continued national administrations but do not speak to problems of more frequent special administrations for handicapped, military personnel, institutional use, which would result in very substantial additional costs were present services to be maintained. The likely result, then, is a combination of reductions in services and increases in the fees changed to students.

C. The third table is derived from the annual report of 1978 and provides a breakdown of research, development and advisory services income and expenditures. The table reveals that ETS funds account in total for somewhat more than one-third of these expenditures and that its contribution is about evenly divided between research on the one hand and development, advisory, and instructional services on the other. This is very much in keeping with ETS charter obligations to contribute to the fund of knowledge in educational methodology and psychometric theory, to develop services and programs needed by students and educational institutions, and to work to improve the understanding and use of tests and other assessment measures.

Although ETS can and does distinguish between research directly supported by test sponsors, ETS does not classify research into that related to testing programs and other categories. To do so would misrepresent the nature of its research activities since much research related to testing is supported by others, including ETS, beyond that which program sponsors alone support. It is in the very nature of research that applicability to specific purposes of findings, conclusions and knowledge gained cannot be assured at the outset or, sometimes, even when the research is in progress. A project which is quite speculative and/or quite theoretical may yield results that have immediate benefit to testing programs.

Conversely, projects believed initially to be directly related to testing programs may turn out to involve broad behavioral concepts or psychometric issues which have no applicability to operating programs at all. Research projects funded by specific testing programs can be identified, as shown in the latest Program Research Progress Report which is enclosed. However, attempting to classify projects as testing research as opposed to educational research would result in arbitrary categorizations of doubtful utility.

Decisions concerning the allocation of ETS funds to research activities are made in the following manner. An overall research plan is reviewed and approved by the Standing Committee on Research and Development and the ETS Board of Trustees. The membership of this standing committee includes both ETS Trustees and other eminent educators and psychometricians. The research plan of the Committee on Research and Development is presented to the Finance Committee of the Board of Trustees which recommends the level of resources to be allocated in support of Research and Development, for approval by the full Board.

Once the overall plan and total resource allocation has been established by the Board of Trustees, responsibility for administration in accordance with research and development objectives rests with ETS management. Operating under the purview of the Standing Committee on Research and Development the Senior Vice President for Research and Development approves an annual plan for allocating ETS support funds to various programs of research and, working with officers and staff, determines which specific projects will be supported. This plan for allocating support, and progress of the specific activities that have been conducted is monitored against the research plan, and reports are provided regularly to the President, other senior ETS officers, and to the Standing Committee on Research and Development.

In addition, ETS annually submits over 100 proposals to federal goverment agencies, foundations and other funding organizations. The vast majority of these are competitive proposals, submitted in response to published requests for proposals or competitions announced by these agencies. As indicated in tables accompanying this letter about two-thirds of ETS research is supported by external agencies, and about one third from ETS funds as outlined above.

We appreciate this opportunity to respond to questions raised during the hearing. Let me also comment on the courteous and professional manner of the members of the subcommittee staff. My staff and I have been treated most cordially during the several months of activity on the proposed bills, H.R. 4949 and H.R. 3564.

Sincerely,

Alice Irby

Alice J. Irby
Vice President

AJI/pm
attachments

Editor's note: The tables referred to in this letter are identical to tables II, III, and IV attached to Ms. Irby's letter dated November 28, which follows.

TABLE I

COLLEGE ENTRANCE EXAMINATION BOARD

Statement of Revenues, Expenses and Fund Balance
Year Ended June 30, 1979
With Comparative Figures for 1977-78
and for 1979-80 Current Projections

	Actual 1977-78*	Actual 1978-79	Current Projections 1979-80
Revenues			
Fees for Programs and Services			
- Admissions Testing Program	$19,205,988	$21,478,660	$22,443,000
- College Scholarship Service	15,267,308	18,819,291	20,332,000
- Advanced Placement Program	3,719,224	4,228,966	5,008,000
- PSAT/NMSQT	3,423,228	3,719,620	4,008,000
- College-Level Examination Program	3,582,263	3,677,820	4,020,000
- Student Search Service	2,134,264	2,574,882	2,767,000
- Puerto Rico Programs	791,753	1,017,589	889,000
- Descriptive Tests	35,570	704,109	478,000
- Comparative Guidance and Placement Program	340,103	304,573	377,000
- Other	1,714,551	1,612,631	1,818,000
Grants and Contracts for Special Services	1,559,774	2,248,870	1,713,000
Membership dues and Meetings	425,697	456,909	507,000
Investment Income	218,702	187,903	225,000
Other	34,896	53,575	108,000
	$52,453,321	$61,085,398	$64,693,000
Expenses			
Programs and services:			
Direct:			
- Admissions Testing Program	$14,964,491	$17,553,540	$17,790,000
- College Scholarship Service	14,112,686	17,248,229	19,576,000
- Advanced Placement Program	3,055,702	3,630,054	4,215,000
- PSAT/NMSQT	1,595,835	1,838,045	1,904,000
- College-Level Examination Program	2,724,071	3,198,548	3,074,000
- Student Search Service	1,007,077	1,141,438	1,412,000
- Puerto Rico Programs	881,140	885,981	1,053,000
- Descriptive Tests	489,107	812,270	752,000
- Comparative Guidance and Placement Program	304,682	264,651	303,000
- Research and Development	1,589,339	2,314,852	2,675,000
- Educational Public Service	1,019,846	878,879	798,000
- Other	896,282	558,430	575,000
Unallocated			
- Regional Office Operations	3,189,223	3,524,511	3,943,000
- Publications	2,111,136	2,317,179	2,636,000
- Other	850,145	1,091,362	1,192,000
Membership Services	546,399	618,753	705,000
General Administration	2,511,100	2,920,181	2,988,000
	$51,848,261	$60,796,903	$65,591,000
Excess of Revenues over Expenses	$ 605,060	$ 288,495	$ (898,000)
Fund Balances			
Beginning of Year	$ 9,787,500	$10,392,560	$10,681,055
End of Year	$10,392,560	$10,681,055	$ 9,783,055

*Certain items in 1977-78 have been reclassified to conform to the 1978-79 presentation.

( ) denotes red figure

October 1, 1979

TABLE II

Educational Testing Service (ETS) Revenues 1959-1979 (10-year intervals)*

(Dollars in Millions)

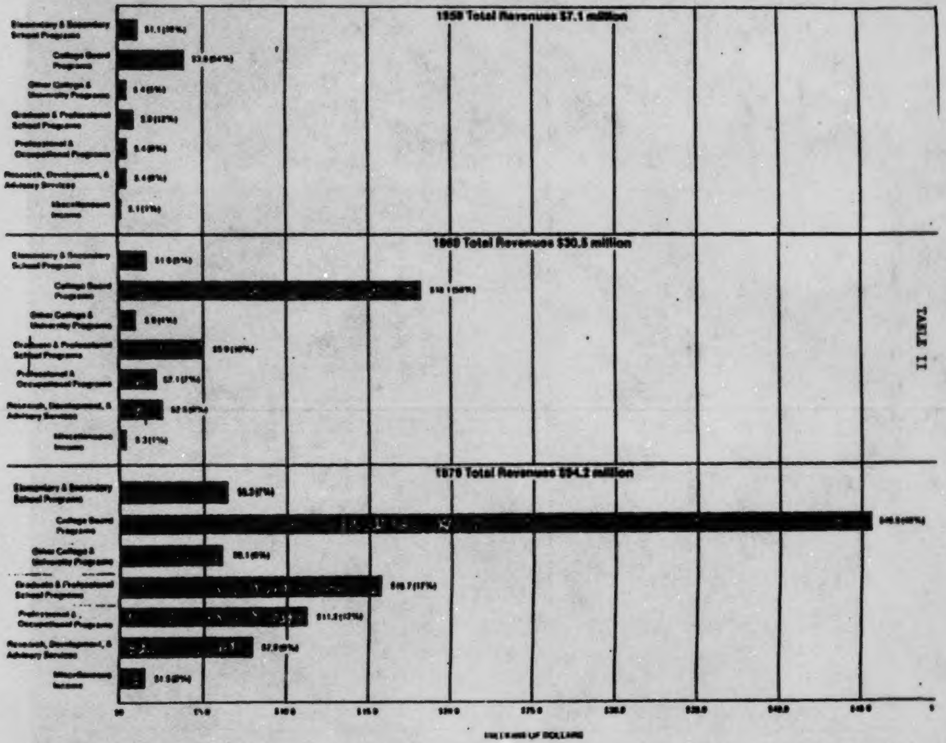


TABLE III

TEST DEVELOPMENT COSTS: SELECTED TESTS

	Total National Adminis-trations	Number of New Forms Now Created Annually	Number of Additional New forms to Keep Current Pattern	Approximate Cost per New Test Form	Total Additional Cost Annually	Number of Test Takers	Current 1979-80 Test Fee
GMAT	8	2	6	$92,000	$552,000	185,000	$12.50
GRE anced Tests	10 Adminis-trations for each of 6 tests = 60	1 for each or 6 total	9 for each of 6 tests = 54	$50,000	$2,700,000	70,000 for 6 Tests	$14.00

Note: Costs for creating a new GRE Advanced Test in any discipline are approximately $50,000 for each new version, and the costs of each version of the GMAT is approximately $92,000. As can be seen from the Table above, the costs of creating enough additional forms of these two tests to permit the current national testing pattern using "secure" forms would approach $2,700,000 annually for the six GRE Advanced Tests only and over $500,000 for the GMAT.

Source: Attachment #6, Mr. Solomon's testimony, July 31, 1979

TABLE IV

RESEARCH, DEVELOPMENT, ADVISORY & INSTRUCTIONAL ACTIVITIES 1978

	Revenue	Expenditures	ETS Contribution
R&D,A&I	$ 6,860,671	$11,038,918	$4,178,247
* - Research only	4,167,000	6,195,000* (See Table III a below)	2,028,000
Development, A&I	2,693,671	4,843,918	2,150,248

TABLE III a

Sources of ETS Research Expenditures - $6,195,000

	Amount	%
Federal Government	$2,636,000	42.6
ETS	2,028,000	32.7
Foundation	477,000	7.7
State & Local Governments	238,000	3.8
Other Sources	816,000	13.2
	$6,195,000	100.0

Source: ETS Annual Report 1978

EDUCATIONAL TESTING SERVICE ETS WASHINGTON, D.C. OFFICE

Suite 300, 1800 Massachusetts Ave., N.W.
Washington, D.C. 20036
•
202-659-0616

November 28, 1979

The Honorable Carl D. Perkins
Chairman
Committee on Education and Labor
Subcommittee on Elementary, Secondary,
and Vocational Education
B-346C Rayburn House Office Building
Washington, D.C. 20515

Dear Chairman Perkins:

As a result of conversations with your staff concerning your letter of November 9, 1979 to Mr. Solomon, his response of November 14, and my letter to the Subcommittee of November 13, 1979, it seems wise to try to arrange the information submitted previously from time to time in a form that is directly responsive to the four points enumerated in your November 9 letter.

Since the four points deal with matters of cost and finances, let me state at the outset that ETS publishes and distributes routinely financial reports concerning its financial affairs which are prepared in accordance with generally accepted accounting standards and certified by independent public accountants. A copy of the 1978 Annual Report is attached. Also enclosed is a document prepared expressly for the Subcommittee entitled Educational Testing Service: Finances ... Where the Money Comes From and Where It Goes, which is an attempt to explain some of the data in the Annual Report and to offer additional data concerning questions raised during the hearings on H.R. 4949 and H.R. 3564. It contains the most recent data for 1978-79.

1. Your letter makes reference to the ETS Project Operating Statement. The statement is an internal document used to monitor income and expenses during the course of a fiscal year. Data are organized by management areas (e.g. by responsibility of Vice Presidents); they are unaudited and require adjustments by cost centers and for overhead charges. Data change monthly and thus, unlike an audited balance sheet and income statement, do not give a full and reliable snapshot of the financial conditions of programs and sponsors. The document also contains information on balances maintained for clients, such balances showing wide cyclical variations during the year reflecting patterns of program activity. We believe that it is inappropriate to circulate documents such as the Project Operating Statement for two reasons: a) Data reflecting revenues, costs and balances for ETS' clients are the property of these independent clients, who have the authority to release such information, and b) Making public financial details of ETS' pricing and cost recovery policies would give competitors considerable and unfair advantage in bidding for contracts.

2. The second point is related to the first in that it requests cost and revenue data by program or individual activity. Providing such information would not only provide an incomplete financial picture of each program but would also be misleading as to full program costs. The programs affected by H.R. 4949 are conducted under contract with independent sponsoring groups. These sponsors have expenses which must be met in addition to ETS's charges. As indicated above, data reflecting total revenues and costs are the property of these independent clients. Often, ETS does not have access to these complete data. Even when it does, it does not have the authority to decide to whom it will be reported. The sponsors make regular financial reports to their memberships and to various governmental agencies. A sample of such a statement is enclosed, labeled Table I. It is a statement of Revenues, Expenses and Fund Balance published by the College Entrance Examination Board on October 1, 1979.

The second table gives a picture of ETS revenues at ten-year intervals from 1959 to 1979. From it, one can ascertain the growth in categories of ETS programs and the changes over time in the composition of total revenues by program area. Two points are of interest:

a. College Board programs, while the largest single component, have declined proportionately since 1959. The $45.5 million in 1979 includes all College Board programs of which the Preliminary Scholastic Aptitude Test and the Admissions Testing Program (SAT and Achievement Tests) are only two.

b. Professional and Occupational Programs have doubled as a proportion of the total, and Graduate and Professional Programs grew from 12% to 17% of the total.

Your second point makes reference to the allocation of "surplus". As stated on page 3 of the enclosure on ETS' Finances, fees charged beyond costs to program clients (and on some contract research as well) average 9-10%. Last year (1978-79), this amounted to $7.5 million. Of this total $6.3 million went to support research, development, advisory and instructional services. The remaining $1.1 million (ETS' net revenues) was used for additions to operating buildings, capital equipment and working capital. Research activities are discussed under item 4 below.

During the debate on H.R. 4949, there were sometimes references to a "profit margin of 22-27%." Such statements, we believe, stemmed from a misunderstanding of financial income and expense reports of the College Board and of ETS. The "22-27%" figure may have come from taking the excess of revenue over direct expenses for the Admissions Testing Program in 1977-78 (see attached nationally distributed College Board Statement of Revenues, Expenses, and Fund Balance, Year Ended June 30, 1979), and using the result as a percentage of revenue:

$19.2 million (revenue)
-14.9 million (expenses)
4.3 million, which is 22.3% of $19.2 million.

Such interpretation, however, ignores a total of $11.1 million in College Board expense for such support services as regional office operations, publications, research and development, administrative costs, support for special conferences, subsidy for test administrations for special populations (such as religious minorities and the handicapped).

Since the Admissions Testing Program accounts for 36.6% of College Board revenues, allocating that percentage of support service expense ($4.085 million) to Admissions Testing Program expense, would bring that program's total expense to $19,050,380 - a 0.8% "surplus" of $155,608 of revenue over total expenses.

In addition to concerns about the proprietary interests of clients, ETS' reluctance to release more detailed cost information than contained herein stems from the fact that it submits proposals in competition with many other organizations, both non-profit and profit, to offer testing, research and related services. Contractual arrangements with each client are determined through negotiation and depend upon the specific circumstances of each situation. Factors which must be considered include the nature of the service to be provided, the transferability of the methods used in designing services for the same or other clients, the opportunities for research or exploration of new methodology, the length of the contract and the financial stability of the client. Sometimes, ETS enters into contracts with a very low fee or none at all to encourage development of a service which meets a pressing social or educational need. As stated above, making public the financial details of each individual activity would give competitors considerable and unfair advantage in bidding for contracts and substantially restrict ETS' ability to negotiate contracts appropriate to particular circumstances. To our knowledge, the Congress does not generally require corporations, non-profit or profit, to release specific data from which cost rates and confidential client data can be derived. To do so in this case would be contrary to customary practice.

3. The third item asks for additional costs, test by test, that would be incurred if H.R. 4949 were enacted. Such projections are not easily derived, for they would depend on a host of decisions to be made by the various sponsors concerning all aspects of compliance with provisions of the legislation. ETS estimates costs of programs from specifications supplied by its clients. For example, the sponsor must specify the number of administrations by type (e.g. national, sabbath, handicapped), the number of new test forms to be created during a given time period, test candidate volume estimates by test and test administration, the number and kinds of publications to accompany the test materials, score reporting services, the kinds of score interpretation materials to be reported as well as information on numerous additional variables. Thus, without such specifications from sponsors as to the kinds and availability of services, ETS cannot project the financial impact of any specific legislation.

We can provide some rough estimates of test development costs, given certain assumptions. The third table is reproduced from Attachment 6 of Mr. Solomon's testimony. It gives examples of projected increases in test development costs, given certain assumptions about disclosure of tests. From the table, one can estimate increases in costs and consequent likely increases in student fees, given different assumptions about the number of new forms produced annually and the number of test takers.

a. The current average cost per test form in GMAT is approximately $92,000. Actual costs vary a bit from year to year depending upon the scheduling of work. GMAT introduces two forms per year. Thus, the production of six new forms (assuming eight national administrations are continued) would add $552,000 annually to a current base of $184,000 or an increase of 300%. If the number of test takers remains constant, the increase in fees would be about $2.00 from test development cost increases alone. This, of course, does not include other administrative costs, such as printing more tests in smaller quantities per test, mailing of forms to students, copying of answer sheets, all of which might lead to increases in fees.

b. The same calculations can be made for the six GRE advanced tests affected by H.R. 4949. Currently test development costs for these exams approximate $300,000 per year. Assuming they are administered ten times in national administrations, the test development cost increase would approximate $2,700,000 or 900%. If the number of test takers remains 80,000, fee increases of $38-$39 could result.

These data assume continued national administrations but do not speak to problems of more frequent special administrations for handicapped, military personnel, institutional use, which would result in very substantial additional costs were present services to be maintained. The likely result, then is a combination of reductions in services and increases in the fees charged to students.

Another way of examining test development costs is to select a state, such as New York, in which either the SAT and/or one or more Achievement tests of the College Board are administered a total of 147 times. Last year the SAT was given 21 times; the 15 Achievement tests were administered several times, some as many as 10, some as few as 4. Presently, ETS prepares 24 new test forms per year. However, of the grand total of 147 occasions for which test forms are administered, 123 depend on the storehouse of secure tests produced in previous years. Because they are secure--because their contents have not been made public--they can be reused. The cost of producing each of the 24 new tests varies, but a very conservative estimate of the cost of developing these 24 would be about $1 million dollars. Now, suppose that because of the requirements of legislation, there are no previously-used secure tests which may be reused. Instead of 24 new forms, the College Board would have to provide 147 new forms to make the tests as readily available as at present.

That is six times the present rate of new test development, for an additional cost per year of $4 million dollars, an increase in test development cost of 400%. It is unlikely that the membership of the College Board would want to pass all of this cost on to the candidates. So the more likely result would be a reduction in the number of tests and the times each test is offered--especially for special administrations to accomodate various special student needs, such as Sunday administrations for religious reasons. Even if legislation would require disclosure of half the test forms, the result would add an additional cost of about $2 million dollars or 200%.

As a general rule of thumb, it is accurate to estimate the cost of test development of a single form of a test to range from $30,000 to $100,000, depending upon the nature of the particular test. In four major programs (College, ATP, GRE, GMAT, LSAT), test development costs currently average 6% of "test-taker" fees. Assuming maintenance of most services, this percentage would increase substantially were H.R. 4949 enacted.

4. Your letter asks for information concerning ETS' research and development expenditures. The fourth table is derived from the annual report of 1978 and provides a breakdown of research, development and advisory services income and expenditures. The table reveals that ETS funds account in total for somewhat more than one-third of these expendirures and that its contribution is about evenly divided between research on the one hand and development, advisory, and instructional services on the other. This is very much in keeping with ETS charter obligations to contribute to the fund of knowledge in educational methodology and psychometric theory, to develop services and programs needed by students and educational institutions, and to work to improve the understanding and use of tests and other assessment measures.

Although ETS can and does distinguish between research directly supported by test sponsors, ETS does not classify research into that related to testing programs and other categories. To do so would misrepresent the nature of its research activities since much research related to testing is supported by others, including ETS, beyond that which program sponsors alone support. It is in the very nature of research that applicability to specific purposes of findings, conclusions and knowledge gained cannot be assured at the outset or, sometimes, even when the research is in progress. A project which is quite speculative and/or quite theoretical may yield results that have immediate benefit to testing programs.

Conversely, projects believed initially to be directly related to testing programs may turn out to involve broad behavioral concepts of psychometric issues which have no applicability to operating programs at all. Research projects funded by specific testing programs can be identified, as shown in the latest Program Research Progress Report which is enclosed. However, attempting to classify projects as testing research as opposed to educational research would result in arbitrary categorizations of doubtful utility.

Decisions concerning the allocation of ETS funds to research activities are made in the following manner. An overall research plan is reviewed and approved by the Standing Committee on Research and Development and the ETS Board of Trustees. The membership of this standing committee includes both ETS Trustees and other eminent educators and psychometricians. The research plan of the Committee on Research and Development is presented to the Finance Committee of the Board of Trustees which recommends the level of resources to be allocated in support of Research and Development, for approval by the full Board.

Once the overall plan and total allocation has been established by the Board of Trustees, responsibility for administration in accordance with research and development objectives rests with ETS management. Operating under the purview of the Standing Committee on Research and Development the Senior Vice President for Research and Development approves an annual plan for allocating ETS support funds to various programs of research and, working with officers and staff, determines which specific projects will be supported. This plan for allocating support, and progress of the specific activities that have been conducted is monitored against the research plan, and reports are provided regularly to the President, other senior ETS officers, and to the Standing Committee on Research and Development.

In addition, ETS annually submits over 100 proposals to federal government agencies, foundations and other funding organizations. The vast majority of these are competitive proposals, submitted in response to published requests for proposals or competitions announced by these agencies. As indicated in tables accompanying this letter about two-thirds of ETS research is supported by external agencies, and about one-third from ETS funds as outlined above.

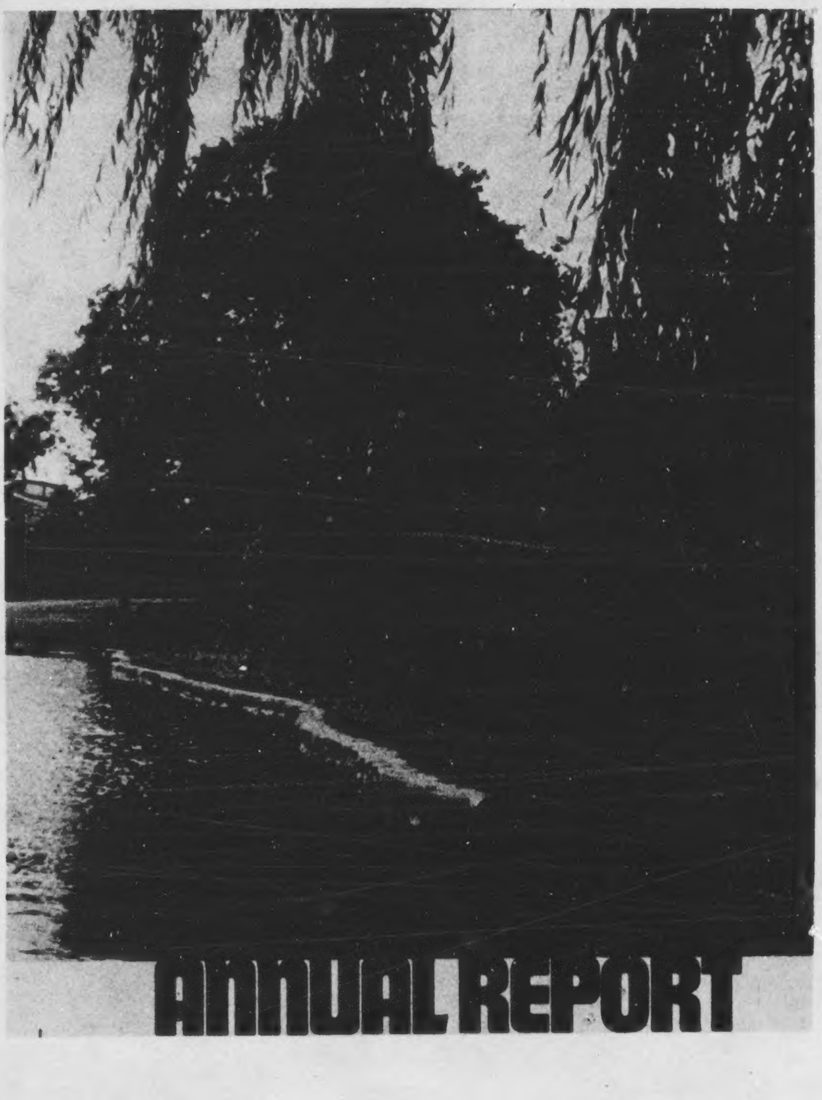
We appreciate this opportunity to elaborate the information supplied previously and to explain our practices regarding our research and development efforts. We appreciate, too, your thoughtful consideration of the issues surrounding the proposed legislation and the professional and courteous manner with which your staff have dealt with those of us who have been associated with them.

Sincerely,

Alice Irby

Alice J. Irby
Vice President

AJI/pm



ANNUAL REPORT

Educational Testing Service (ETS) is a private, nonprofit organization devoted to measurement and research, primarily in the field of education. It was founded in 1947 by the American Council on Education, the Carnegie Foundation for the Advancement of Teaching, and the College Entrance Examination Board. The presidents of the American Council on Education and the College Board serve *ex officio* on the ETS Board of Trustees, as does the president of ETS. Thirteen other members are selected from the education and business communities to serve four- year terms. The main office of ETS is in Princeton, New Jersey, and there are nine regional offices.

ETS, A NONPROFIT ORGANIZATION, IS AN EQUAL OPPORTUNITY EMPLOYER.

COVER: Across the pond, a glimpse of the Henry Chauncey Conference Center at ETS in Princeton. Photographed by Del Merritt, Visual Concepts Unlimited, Trenton, NJ.

COPYRIGHT © 1978 BY EDUCATIONAL TESTING SERVICE.
LIBRARY OF CONGRESS CATALOG CARD NUMBER 50-30908.

ETS 1978 Annual Report

2 The President's Report:
Achievement Test Scores in Perspective
5 Programs & Services
11 Auditors' Report & Financial Statements
14 Five-Year Summary of Revenue
15 Research & Development
19 Staff & Administration
21 Some Current ETS Publications
22 A Partial List of ETS Clients, 1977-78
24 ETS Trustees, 1977-78

Educational Testing Service • Princeton, New Jersey 08541

The President's Report:

Achievement Test Scores in Perspective

The need to assess how much students have learned has been fundamental in education for as long as there have been students and teachers. Long before standardized tests of achievement came on the scene, teachers were making such judgments. They based them on information gleaned from familiar sources: direct observation of students' work, class recitations, conversations with the student or other teachers, daily quizzes, and final examinations. All these bits of information entered into the grades the teacher awarded. They still do—and they should.

Since we had all these techniques at our disposal in Horace Mann's day and long before, why were standardized achievement tests welcomed when they came along? There are several reasons.

An Inexact Business

Teachers knew what an inexact business grading really is, and welcomed a new development that held promise of improving their information. Among the virtues of standardized tests, three were particularly appealing. The first virtue was *accuracy*. Studies were made to compare the amount of random error in the traditional kinds of information with the amount of error in the newer standardized tests in the same subjects, and the findings were consistent: the test scores were consistently more accurate. Furthermore, they had a second, related virtue—*objectivity*—that helped overcome some other problems. Some teachers marked hard and some marked easy, some based their grades heavily on deportment, some on how hard students tried rather than how much they achieved. The standardized tests knew nothing of these things, nor of sex, race, honesty, amiability, or love of one's fellow man: only how much the student had learned. Third, scores on standardized tests of achievement had the virtue of *comparability:* They showed not only how well particular students performed in comparison with their classmates but also how well they did in comparison with pupils in other classes, other schools, and other states.

We had, then, the basis for a fine combination of techniques: standardized tests, which could measure sheer accomplishment in several areas very well, and teacher judgment, which could add dimensions inaccessible to standardized testing but pertinent to the interpretation of pupil scores.

We have learned a hard lesson in the ensuing years: how difficult it is to keep achievement scores in that perspective, to see them as a valuable ingredient in the mix of information that indicates how much an individual student or a group of students has learned in a particular area. Scores are not, by themselves, sufficient. That fact needs to be reiterated. But they do remain a key component in any adequate learning information system.

Three Fallacies

Undoubtedly, testing has suffered more from the excessive expectations of its most devoted advocates than from the attacks of its critics. I have mentioned three virtues of tests that have enhanced their usefulness. Let me balance the account by mentioning three fallacies, all born of overenthusiasm, that have encouraged the misuse of test scores.

testing has
suffered from
the excessive
expectations of its
most devoted
advocates

The first I shall call the *Micrometer Fallacy.* Some people have invested test scores with a precision—an infallibility—that they never possessed. The ensuing discovery that they are not perfectly accurate has obscured the fact that they are more accurate than most of the alternatives.

The second error I shall call the *Whole Person Fallacy:* the tendency to read into achievement test scores much more than they really tell, which is simply the amount a student has learned in a given subject. Some people are let down upon discovering that achievement tests fail to measure a variety of traits like honesty or leadership or social consciousness, and in their disappointment over the fact that the tests do not describe the whole person, forget that they do a rather good job of measuring the academic accomplishments they purport to measure.

Third, there is the *Equal Preparation Fallacy.* Some people expect the test to compensate somehow for the differences in academic development of children whose learning opportunities have differed dramatically. The test score tells you nothing about the difficulties a student has had to overcome to acquire a given level of proficiency, but it tells you a great deal about what that level is—a fact that is central to deciding what the student is ready to tackle next.

Comparable Results

The same three virtues of test scores—accuracy, objectivity, and comparability—that had appealed to teachers interested in the accomplishments of individual students also commended themselves to administrators concerned with the learning of groups of students. For the administrators the greatest of the virtues was (and is) comparability.

They saw at once the power of standardized tests to permit comparisons of the learning achieved by pupils of different teachers, or in different schools, or in different districts. Since it was well known that some schools were much more demanding than others in their curriculum and in their grades, here at last was a chance to put all students on the same footing in an objective comparison of results across schools—to make sure the children were learning as much as they should, as determined by how well the children elsewhere were doing.

> “The test score tells you nothing about the difficulties a student has had to overcome . . . ”

In the main, these aspirations were sound. And when the three fallacies have been resisted—when people have not expected the precision of the micrometer, have not looked to achievement tests to measure the whole person, and have not assumed that by standardizing the test you have standardized preparation—then, indeed, the scores have given a new dimension, through their comparability across geographic areas and spans of time, to the information available to educators. It is hard to imagine any other basis we might have had for learning, for example, that the verbal and mathematical skills of students applying to college have declined in the last 15 years. It may not have been a comforting message, but it was one worth getting.

Let's look at a third major use of standardized achievement test scores—in the selection of students, especially for college. This use got an important boost during World War II, when teacher power was unavailable to grade the College Board essay tests. The Board substituted objective tests as a war measure because they could be graded so efficiently, fully expecting to return to the essays after the war. But to the surprise of many, not only were the objective tests easier to grade, the scores on them were at least as effective as the essay test scores had been in measuring the attainments of the students who took them. This conclusion was arrived at first through the general observation of the school and college people involved and then confirmed through careful research. The scores were by and large accurate—more so than scores on the essay tests had been.

Students who were admitted proved to have the accomplishments the test scores had promised. The scores provided a common currency, unlike the grade-point averages that reflected differences in grading between schools and between parts of the country. Moreover, they were much less subject to special coaching or "prepping" than the essay tests had been; they reflected more fairly the accomplishments of students from different parts of the country, different schools, different curricula.

Unparalleled Growth

As a result, the selective institutions that used the test scores as one basis for admissions decisions were able to seek out talented students from every corner of the nation and every social stratum. They began to see on campus a more heterogeneous group of students both socially and geographically. Moreover, with objective tests of aptitude and achievement in place and efficient, it was possible to test the enormous postwar wave of college applicants under the new system, and an unparalleled expansion of higher education took place in this country.

The system worked because of the same virtues of standardized tests: accuracy, objectivity, and comparability. Again, however, some people were tempted into entertaining great expectations that could not be and have not been fulfilled. The scores were accurate in the main but not micrometers in their precision, even though their errors were smaller than those of other known techniques. They measured achievement and readiness, but not traits of character and temperament—not the whole person. They made no allowance for the inadequacies of preparation or special circumstances of environment that the student had overcome in school. In short, the assumption of equal preparation remained—and remains—a fallacy.

Guidance for Students

We have discussed three legitimate and important uses of standardized tests: their use by teachers to determine how much the individual student has learned, by administrators to determine how much classes and larger groups have learned, and by university admissions people to discover how well prepared a prospective student may be. Each use has made great contributions to education and society when kept in perspective and used with other pertinent information.

Another important use of test scores is by students—to help them as they examine their educational and career goals and estimate their readiness to undertake a more or less demanding program of study as their next step. This use has received less emphasis historically than it should have, but the evidence from a set of tests can add materially to the information available for a student's guidance.

People who now discover that test scores can vary from day to day and test to test are on the right track. Every examination, every judgment about people, is fallible and has a typical error rate. The standard error of measurement associated with standardized test scores is well known because it is readily determined and regularly announced by the publishers. That does not mean that other forms of measurement, such as essays or interviews or letters of reference or teachers' grades, have no error or a smaller error. In fact, although such errors are rarely reported or even determined, research indicates that, in the typical case, they are much greater than the standard error of measurement of a test score.

Those who call for a high degree of teacher involvement in assessing students in their classes are right, as are those who decry the use of test scores alone as a basis for evaluating the effectiveness of an entire educational program. There is much more to be considered than is reflected in the score, including the conditions under which the results were achieved. But to say what some recent critics have said—that even if one wants to know how pupils are doing in arithmetic one should be forbidden to give a standardized test in arithmetic—is to swing the pendulum right out of the clock.

Colleges and universities that use test scores in selective admissions should indeed use other information as well. Achievement tests simply cannot measure the whole person. Obviously, a college should consider the record of previous school performance and the judgments of teachers and counselors. Yet there are many gifted youngsters who make mediocre records in school but surface through their test scores, as Julian Stanley's work at Johns Hopkins University has amply demonstrated.

> ‘. . . there are many gifted youngsters who make mediocre records in school but surface through their test scores . . . ’

Forget the Magic

Finally, those who would ascribe to tests given at school age some magic that enables them to divine genetic intelligence or ability to learn should forget it. Achievement tests measure developed ability—developed in relation to a particular subject or discipline. To a large extent, so do scholastic aptitude tests, although the areas of experience through which a student is prepared for them are much more long term and more pervasive in our society. But equal opportunity in education simply has not been realized. To think that, at 18 years of age, people whose experiences have been vastly different can show their inborn potential through a test of verbal and mathematical reasoning is naive, regardless of their cultural advantages or disadvantages.

Standardized tests of achievement have amply demonstrated their utility over the past two or three generations. Because of their accuracy, their objectivity, and their comparability, they deserve recognition as a powerful tool in education. They have suffered in esteem first at the hands of those who claimed for them a set of qualities they could never attain, and latterly from the protests of those who have proposed that since they are imperfect they be done away with. I suggest we put them in a reasonable perspective as we strive to improve both the tests and their use.

W W Turnbull

Programs & Services

When Educational Testing Service was established in 1948, its explicit purposes were to develop responsible testing programs and related services, to conduct educational research, and to provide advisory and instructional services in measurement. Implicitly, the new organization was expected to reach out to the educational community, to help analyze and meet its needs.

In 1977-78, the year of its thirtieth anniversary, ETS could perhaps be characterized best as "reaching out."

In the Field

Through a network of regional offices, the organization provides a variety of services for educators and test-takers in the field. Launched in 1948 with a modest office in Los Angeles, regional operations expanded gradually and now include eight offices spanning the continent, plus one in Puerto Rico. This network enables the organization to respond swiftly to requests, wherever they originate, and generates a steady flow of information to the main office in Princeton about education and measurement needs in all parts of the nation.

Regional staff members regularly interact with educators, students, parents, and others with an active interest in education and measurement. For instance, the Southern, Midwestern, and Northeastern offices are Technical Assistance Centers (TACS) designated by the U.S. Office of Education to help state and local school officials in their regions evaluate programs for the disadvantaged under Title I of the Elementary and Secondary Education Act (ESEA). In this role, they provide consultation to state departments of education, conduct workshops for Title I evaluators and classroom teachers, meet with advisory committees composed of parents, and undertake various other tasks.

Community relations specialist Katherine Moore leads visitors on a tour of ETS in Princeton.

At the largest regional office, in Berkeley, there is an extensive array of activities, including testing program services, consultation, and research. Among the projects under way in 1977-78: a national assessment of the need for educational media and materials to teach the handicapped, and a survey of the ways universities and colleges have adapted to the changing conditions of the 1970s.

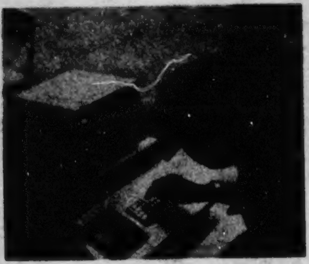
In Wellesley Hills (Boston), staff studied the measurement aspects of performance-based high school graduation requirements. In Evanston, a process was developed for identifying and placing academically gifted children. Los Angeles staff continued to work with various school districts in the evaluation of bilingual programs. Meanwhile, plans were drawn to expand the Austin office so it could provide a broader array of services to the Southwestern states.

One pioneering development was the creation of a Regional Test Development Center at the Southern office to serve the employment testing needs of a consortium of eight Southeastern states. The center grew out of a 1977 study by ETS staff in Atlanta and at the Center for Occupational and Professional Assessment (COPA) in Princeton. They explored the feasibility of a multistate approach to developing and validating tests for the selection of state employees. The study was funded by the U.S. Civil Service Commission and conducted under a contract with the State of Georgia.

International Activities

For many years, ETS has engaged in measurement-related activities around the world. Its *International Newsletter* is a vehicle for the exchange of information by some 2,500 psychologists and educators in 110 nations, and its annual summer workshops in measurement for foreign scholars have attracted more than 400 participants in 17 years.

Several ETS-developed tests are administered in faraway places each year, including the College Board's Scholastic Aptitude Test (SAT), the Graduate Record Examinations (GRE), the Graduate Management Admission Test (GMAT), the Law School Admission Test (LSAT), and the Test of English as a Foreign Language (TOEFL). Sponsored jointly by the College Board and the GRE Board, TOEFL is designed especially to measure the English language skills of the thousands of foreign students who annually seek admission to U.S. universities and colleges. In 1977-78, TOEFL was administered to a record 192,000 at home and abroad.



Left: Foreign scholars do computer terminal exercise in the cost effectiveness of an educational program. Right: *They are, from left, Chang Lok Chin, Malaysia, and Dossi Natarajan, Singapore.*

During the year, ETS continued working with the 14-nation Caribbean Examinations Council to develop the council's capability for educational testing and related data processing. Started in 1973, this cooperative effort has produced tests appropriate to the region that will begin, in 1979, to supplant the traditional school-leaving examinations from the United Kingdom. As the council gains experience and expertise, the role of ETS will be reduced gradually, then phased out.

Active participation in the International Association for Educational Assessment (IAEA), which ETS was instrumental in establishing in 1976, continued in 1977-78. For IAEA, ETS began to study the feasibility of developing a multinational aptitude test to facilitate college admissions across national boundaries. The test would be administered in a student's native language, with a separate examination determining skill in the language of the country where he or she plans to study. Thus, language problems would not influence aptitude test scores.

Members of the ETS staff visited several nations during the year to consult with education officials and conduct research. One consultant advised an arm of the Iranian Ministry of Education regarding methods of administering and processing that nation's college entrance examinations. Others visited Indonesia and countries in Africa and Latin America in a continuing study of nonformal education—out-of-school learning related to job skills and living standards—in developing countries. This study is funded by the U.S. Agency for International Development.

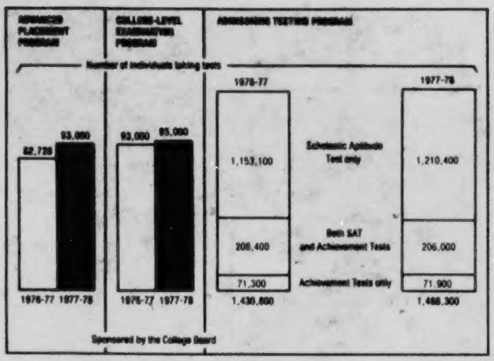
Continuing Education

Workshops conducted by ETS during the year covered a lot of ground, both geographically and topically. Developed by the Programs of Continuing Education staff in Princeton, the workshops were offered both at that site and around the nation, often conducted by regional office staff. Topics ranged from assessing minimum competence in the basic skills to evaluating programs for gifted students, and also encompassed the legal issues involved in teacher evaluation, methods of improving instruction, program evaluation, testing minority groups, values education, and aspects of measurement.

Throughout the year, ETS continued to conduct Evaluation Improvement Program (EIP) workshops across the nation. Based on materials developed by the California Department of Education and published by ETS, EIP is intended to help school personnel become self-sufficient in evaluating educational programs and to increase their awareness of available technical assistance.

With the National Council on Measurement in Education (NCME), ETS presented a nationwide series of seminars on issues in testing for school board members and administrators. Each seminar was tailored to the needs of local participants as they dealt with students, parents, and their communities at large. Among the topics: the assessment of writing skills, how to communicate test results to parents and other citizens, and what teachers can learn from results of criterion-referenced tests.

Arturo Madrid, center, University of Minnesota, confers with Jorge Barovero and Carmen Godinez Windhorst at ETS in Berkeley.



Student Services

In its first full year of operation, the ETS Office for Minority Education published several monographs and a compendium of current ETS services and projects relating to the improvement of educational opportunity for minority groups. *College Achievement through Self-Help*, a manual for minority students enrolled or intending to enroll in college, provided analyses of the common problems they face in adjusting to campus and classroom life and offered guidance in the handling of each problem.

In January 1978, the new Financial Aid Form (FAF) became the sole document for students to file with the College Board's College Scholarship Service (CSS) in order to apply for assistance in meeting college costs. Phased in over a two-and-a-half-year period, the FAF is a standard application form used by universities and colleges across the nation, aid and scholarship agencies in 30 states, and the federal Basic Educational Opportunity Grants program.

Each FAF is processed by CSS and forwarded to the institutions and agencies designated by the applicant. Use of the standard form resulted in a substantial increase in the number of applications filed with CSS.

During the year, CSS also initiated the Early Financial Aid Planning Service to help high school juniors (and younger students) and their families calculate in advance what they can afford to spend on college, how much aid they will need to attend the institutions of their choice, and what aid programs are available to them.

Testing Programs

Continued nationwide concern with educational standards was reflected in numerous ETS activities during 1977-78. It was the first year that the Basic Skills Assessment program was used across the country to help local school districts determine whether secondary school students could read, write, and compute well enough to meet the demands of adult life or the levels of competence required for high school graduation, as measured against locally established standards.

Developed by ETS in cooperation with a consortium of public school districts, the Basic Skills Assessment tests were administered to more than 70,000 students in 200 school districts. ETS provided the tests and related materials and scoring and reporting services, tailoring its reports to the specified needs of the various districts. Staff also provided guidance to district personnel regarding methods of setting appropriate standards.

College Freshmen

Meanwhile, many universities and colleges intensified their efforts to assess the skills of entering freshmen, primarily for purposes of placement. In 1977-78, the California State University and College System administered its new English Placement Test, developed with ETS, to all first-year students—about 25,000 of them—at its 19 institutions. Results were used for student placement and the improvement of curriculum and instruction.

In New Jersey, starting this fall, all freshmen at the state university and colleges will take College Board tests developed by ETS to ascertain whether they are adequately prepared in the basic skills to proceed with higher education or need remedial courses first. Private institutions in the state were invited to participate in the program voluntarily. As in California, test results will be used for both placement and curriculum planning.

The Third 'R'

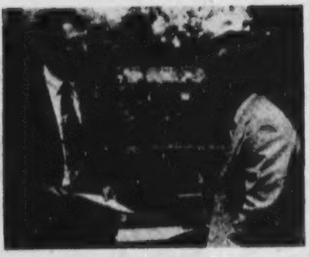
During 1977-78, there was a marked increase in attempts to improve the writing skills of students at all levels of education. At ETS, this trend led to the creation of Programs for the Assessment of

Writing. The center was designed to respond to requests for consultation, training, and related services from state and local education agencies seeking to improve writing instruction and measure the writing abilities of their students.

In its first months, the center received numerous requests to provide inservice training of faculty in the holistic method of scoring essays. Because it dramatically reduces the time required to read and grade compositions, holistic scoring enables English teachers to increase the number of writing assignments they give to their classes without burying themselves in paper work.

Essentially, holistic scoring is based on the assumption that the total impression made by a piece of writing serves as a reliable measure of the author's skills. That is, if a piece communicates effectively, it reflects skill in grammar, syntax, and the other components of writing. So essays are read rapidly for clarity and impact rather than painstakingly for errors. Two or more readers rate each essay on a 1-5 scale, and the total is the student's score. Aside from its usefulness in the classroom, holistic scoring has been instrumental in a resurgence of essay sections in ETS-developed English composition tests.

Evans Alloway, director of Programs for the Assessment of Writing, reflects on his busy schedule.

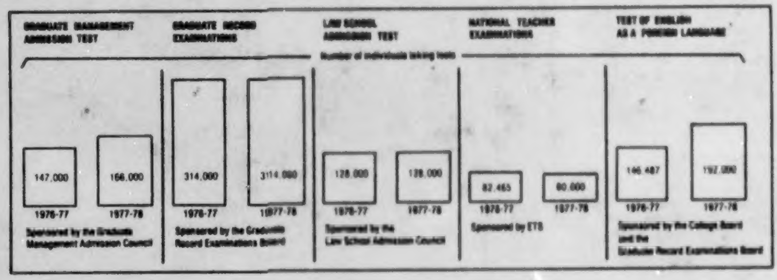


Director Andre Diaz, right, discusses the fast-growing GMAT program with associate director John High.

Sequential Testing

Norming of the completely revised School and College Ability Tests (SCAT) and Sequential Tests of Educational Progress (STEP) was completed in 1977-78 with more than 200,000 students in grades two through 12 participating nationwide. With SCAT-STEP III and the CIRCUS measures for children in preschool and the early grades, ETS now has tests available—for the first time—that permit monitoring a student's progress in major areas of instruction from preschool through the twelfth grade.

During STEP III's development, measurement specialists devised a new technique for relating test results to grade level. ETS does not report grade-equivalent scores on grounds that they are misleading and are based on estimates rather than actual performance data. For example, to say that, on the basis of his or her score on a third-grade arithmetic test, a student has performed at the level expected in the third month of the sixth grade is inaccurate because the performance was not on a sixth-grade test but a third-grade one that may never have been given to any sixth grader.



Under the ETS Grade Level Indicator (GLI) system, a brief "locator test" may be administered to a student to determine which level test he or she should take. If a fourth grader's locator test score is high, he or she will be most appropriately tested on fifth-grade material. STEP results will give an accurate indication of achievement relative to the average performance of all students in a particular grade and month, as measured during the norming process. Used in conjunction with the standard score and national percentile rank reports, the GLI adds a dimension to the performance profile sketched by STEP.

Other Developments

In its second year, the Undergraduate Assessment Program (UAP) offered revised Area Tests in social science, the humanities, and natural science, and Field Tests in business and physical education. About 375 universities and colleges participated in UAP, using its broad range of tests and related services to assess student performance and evaluate curriculums.

Efforts to improve operating systems resulted in speedier processing of score reports following administrations of the Graduate Management Admission Test (GMAT). That program's sponsoring organization, the Graduate Management Admission Council (GMAC), established a committee to plan and facilitate research into graduate management education. GMAC also formulated a policy statement on the appropriate use of GMAT scores in making admissions decisions.

Some developments in other programs during 1977-78: A new Graduate Record Examinations (GRE) section measuring analytic skills was introduced, marking an effort to assess aptitude in other than verbal and quantitative areas; the TOEFL Council embarked on an extensive program of research into various aspects of its test's validity; for the Law School Admission Council, ETS continued development of the Law School Interactive System, a pilot program linking law school admissions offices directly with the ETS computer records of candidates who authorize release of the data.

For the College Board: ETS completed norming of the revised General Examinations in the College-Level Examination Program (CLEP); an essay section was added to the CLEP English Composition Test for administration twice a year; the Descriptive Tests of Language Skills were being field tested, and ETS began developing similar Descriptive Tests of Mathematics Skills to help institutions place students and plan programs.

Left: *Auto mechanic William Y. Johnson tends to business at Cathcart Pontiac, Trenton. Sleeve patches indicate he is certified under a program conducted by* ETS. Right: *Leo Walsed, left, director of the auto mechanics program, examines data with test specialist Edwin Kay.*

Arthur Kroll, director of guidance programs, reviews the Career Skills Assessment Program with its assistant director, Jane Brewer.

Looking at Careers

As the job market changes in this postindustrial society, there is increasing interest in careers. Students are interested in choosing careers and preparing for them. Employers are interested in how to screen applicants, select employees, and assess their competence. The agencies and boards that govern professional certification and occupational licensing are interested in determining the adequacy of education and preparation.

In a variety of ways, ETS attempts to respond to all of these interests.

Career Skills

For the College Board and a consortium of five states—Georgia, Maryland, Minnesota, New Jersey, and Ohio—ETS developed instruments to measure six skills that are essential to career

7

Programs of Vocational Education director Raymond G. Wasdyke peruses new PAYES material.

preparedness: self-evaluation and development, career awareness, decision making, employment seeking, work effectiveness, and personal economics.

More than three years in the making, the result is the Career Skills Assessment Program, available in experimental form during 1978-79. A key feature of the program is the underlying conviction that these career development skills can be not only identified and measured but taught. In fact, one major use of the measures is expected to be the identification of skills that need additional attention in the curriculum.

Guidance

During the year, ETS Programs of Vocational Education completed development of a battery of measures for use in providing career guidance and counseling to disadvantaged adolescents and young adults in Comprehensive Employment and Training Act (CETA) programs, skills training centers, vocational high schools, and correctional institutions. Called Programs for Assessing Youth Employment Skills (PAYES), the measures help identify attitudes, cognitive skills, and social competencies related to employment. They rely heavily on pictorial materials and are administered orally to minimize the impact of limited verbal skill.

Meanwhile, the ETS computerized career guidance program for college students was introduced to several new campuses. Called SIGI (for System of Interactive Guidance and Information), the program is now helping students identify their career interests and values, relate them to various occupations, and progress toward informed and rational career decisions at 14 two-year and 14 four-year colleges.

In the Workplace

In 1977-78, the ETS Center for Occupational and Professional Assessment (COPA) continued to grow, reflecting the trend toward more precise standards of job preparation and qualification. During its first five years, COPA has undertaken numerous measurement and research projects in a wide variety of occupations. Each was designed to assist in the certification, licensing, selection, or self-assessment of people entering or working in a particular field.

COPA's three largest programs illustrate the center's diversity. The Real Estate Licensing Examination program tested 175,000 aspiring brokers and agents in 26 states during the year. The Certified Auto Mechanics Program, sponsored by the National Institute for Automotive Service Excellence, issued certificates to more than 56,000 auto and truck repairmen in a number of mechanical specialties. Some 42,000 would-be insurance agents were tested in seven states under the Multistate Insurance Licensing Program.

New endeavors by COPA's 100-member staff include the evaluation of state training programs for persons who seek certification as applicators of dangerous pesticides. Such certification is required by federal law before a person can purchase the deadly chemicals. Involving interviews and onsite observations before and after training, as well as tests, the project is sponsored by the U.S. Environmental Protection Agency and the Council on Environmental Quality.

Left: *COPA test specialist Suzanne Marder, left, and health programs director Toby Friedman discuss respiratory therapist certification program with its director, Rufus Smith.* Right: *Certified therapist John R. Lloyd works with a patient at Princeton Medical Center.*

Left: *Radiologic technologist Judith White, certified through a COPA program, works at the Princeton Medical Center.* Right: *William D. Hogan, director of that program, discusses operational details with the assistant director, Betty Gronikowski.*

AUDITORS' REPORT & FINANCIAL STATEMENTS

COOPERS & LYBRAND
CERTIFIED PUBLIC ACCOUNTANTS

To the Board of Trustees,
Educational Testing Service:

We have examined the balance sheets of Educational Testing Service as of June 30, 1978 and 1977, and the related statements of revenue, expenditures and fund balance and changes in financial position for the years then ended. Our examinations were made in accordance with generally accepted auditing standards and, accordingly, included such tests of the accounting records and such other auditing procedures as we considered necessary in the circumstances.

In our opinion, the financial statements referred to above present fairly the financial position of Educational Testing Service as of June 30, 1978 and 1977, and the results of its operations and the changes in its financial position for the years then ended, in conformity with generally accepted accounting principles applied on a consistent basis.

Coopers & Lybrand

Newark, New Jersey
August 16, 1978

EDUCATIONAL TESTING SERVICE
BALANCE SHEETS—June 30, 1978 and 1977

ASSETS	1978	1977
Current assets:		
Cash	$ 2,961,873	$ 4,520,062
Short-term investments (approximates market)	5,600,000	4,300,000
Accounts receivable, less allowance for doubtful accounts:		
($100,000 in 1978 and $125,000 in 1977)	3,425,428	[illegible]
Unbilled amounts under contracts	3,367,473	[illegible]
Prepaid expenses	1,408,102	[illegible]
Total current assets	16,862,577	[illegible]
Investments in marketable securities (market value 1978, $4,155,000 including in 1978,		
$828,000 as equipment replacement fund)	4,492,010	4,495,288
Property, plant, furniture and equipment, net (Note 2)	24,031,079	22,774,734
	$45,385,666	$40,945,1[illegible]

LIABILITIES	1978	1977
Current liabilities:		
Accounts payable and accrued expenses	$ 5,938,180	$ [illegible]
Unexpended receipts:		
Under contracts	6,912,917	[illegible]
Under grants including advance payments	1,269,808	1,449,746
Total current liabilities	14,120,905	[illegible]
Deferred income (Note 3)	678,861	
Unexpended receipts (Note 4)	3,809,463	[illegible]
Fund balance (Note 5)	26,771,717	[illegible]
	$45,385,666	[illegible]

See notes to financial statements.

STATEMENTS OF REVENUE, EXPENDITURES AND FUND BALANCES

for the years ended June 30, 1976 and 1977

	1976	1977
[illegible]		
[illegible]	[illegible]	[illegible]
Research, development, advisory and [illegible] services (Note 1)	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]		
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	1,114,[illegible]
[illegible]	1,902,984	[illegible]
Freight and postage	5,411,555	[illegible]
Telephone and telegraph	1,219,568	[illegible]
Other operating and administrative expenses	6,469,266	[illegible]
Total expenditures	[illegible]	[illegible]
Excess of revenue over expenditures	1,356,649	[illegible]
Fund balance, beginning of year	25,417,068	[illegible]
Fund balance, end of year	$26,773,717	[illegible]

See notes to financial statements.

STATEMENTS OF CHANGES IN FINANCIAL POSITION

for the years ended June 30, 1976 and 1977

	1976	1977
Working capital was provided from:		
Operations:		
Excess of revenue over expenditures	$ 1,356,649	$ 901,033
Add items not requiring outlay of funds:		
Depreciation	1,648,371	1,436,670
Allowance for valuation of other assets		10,000
Amortization of deferred income	(76,873)	
Provided from operations	2,928,147	2,347,703
Proceeds from sale of land and building (Note 2)	1,711,896	
Decrease in investments	3,262	
Disposals of equipment		39,736
Increase in unexpended [illegible]	1,032,141	
	5,675,446	2,387,439
Working capital was used for:		
Additions to property, plant, [illegible] and equipment	4,411,977	1,409,177
Increase in investments		1,371,758
Increase in other assets before allowance for valuation		10,000
Decrease in unexpended [illegible]		156,700
	4,411,977	3,857,631
Increase (decrease) in working capital	$ 1,263,469	($ 1,470,192)

See notes to financial statements.

[illegible]

5. *Fund Balance:*

At year end, fund balance is represented by:

	1978	1977
Property, plant, furniture and equipment, less accumulated depreciation	$34,[illegible]	$33,774,7[illegible]
Other, which is fully appropriated as a result of appropriations	2,[illegible]	2,[illegible]
	$[illegible],772,717	$[illegible]

At June 30, 1978, the Corporation had appropriated $2,155,000 for purchases of plant, furniture and equipment and $4,764,000 for research, development, advisory and instructional services.

6. *Research Expenditures:*

The Corporation's research, development, advisory and instructional services expenditures for the years ended June 30, 1978 and 1977 amounted to $11,[illegible] and $11,452,000, respectively, of which $6,860,871 and $7,202,403, respectively, was received from outside contractors and grantors.

7. *Retirement Expenditures:*

The Corporation's retirement plan expenditures for the years ended June 30, 1978 and 1977 were approximately $3,000,000 and $2,847,000, respectively.

8. *Lease Commitments:*

The following is a schedule, by years, of future minimum rental payments required under operating leases that have [illegible] one year as of June 30, 1978.

Year ended June 30:

1979	$ [illegible]
1980	[illegible]
1981	[illegible]
1982	[illegible]
1983	[illegible]
Later years	[illegible]
Total minimum payments required	$[illegible]

Rental expense for all operating leases [illegible] $[illegible] and $1,[illegible] in 1978 and 1977, respectively.

9. *Major Customers:*

During the year ended June 30, 1978, the Corporation's revenues of $38,170,393 [illegible] College Board.

10. *Litigation:*

The Company is a co-defendant [illegible]

[illegible] Educational Testing Service [illegible]

FIVE-YEAR SUMMARY OF REVENUE

	1978	1977	1976	1975	1974
College Board Testing Programs and Services	$38,170,393	$33,168,875	$30,809,828	$27,577,162	$26,776,[illegible]
Graduate and Professional School Testing Programs	14,399,328	13,652,287	12,566,157	11,731,892	11,308,043
Testing of Teachers and Other Professionals	8,606,727	6,577,947	5,807,589	5,087,719	5,288,306
Other Testing Activities at Various Academic Levels	10,333,443	8,261,160	6,688,391	6,123,785	5,659,638
Total Testing Activities	71,508,890	61,660,319	55,870,939	50,520,558	47,025,047
Research, Development, and Instructional Services	6,860,871	7,202,403	5,849,161	7,003,462	6,011,448
Other Income	1,256,782	1,144,124	1,234,250	1,002,208	546,222
Total Revenue	$79,626,543	$70,006,846	$62,954,350	$58,526,228	$53,582,717

Research & Development

Through its research efforts, no less than its programs and services, ETS attempts to reach out to the world around it. With approximately 300 projects under way in 1977-78, the research effort engaged the time and talents of some 200 staff, about half of them with advanced degrees in the social sciences.

In broad terms, they were seeking new knowledge about human learning and development, evaluating educational programs, analyzing educational policies and practices, developing and testing measurement methodologies, and defining and refining statistical data analysis theories and techniques. Their work was supported by grants from and contracts with government agencies, private foundations, organizations that sponsor various testing programs—and by allocations of ETS funds.

Each year since its inception, ETS has channeled a share of its own resources into educational and psychological research. In recent years that share has grown, and in 1977-78, ETS funds accounted for 32.7 percent of the total expended for research (see chart).

Development Goals

In order to coordinate research efforts for maximum benefit to education and society, and to allocate resources wisely, ETS devoted considerable effort during the year to formulating a statement of research and development goals and objectives. In the process, five fundamental goals were defined. They are:

- To improve access to educational opportunity.
- To improve the quality of education.
- To increase the effectiveness of education.
- To enhance understanding of human development and educational processes.
- To enrich understanding of the educational needs and attributes of selected populations.

Research in the areas suggested by these goals is not new to ETS. Numerous projects over the past 30 years have related to each of the five areas. The significance of the goals is not that they point in wholly new directions but that they signal areas of continued emphasis for the future. In 1977-78, several projects were under way in each of the five areas. Some of them are described briefly here.

Access to Opportunity

Recent events have given new impetus to ETS efforts to identify characteristics, in addition to the traditional grade-point averages and test scores, that may be useful in the admissions process—and to develop ways of assessing such criteria.

For years, ETS and sponsors of national testing programs have advised institutions not to place undue reliance on test scores or, indeed, on any single admissions criterion, but to take into account all of an applicant's relevant characteristics. A current project called New Dimensions in Admissions is expected to result in a broader view of such supplementary characteristics and the ways they might be more systematically and reliably assessed.

No prior assumptions have been made about the nature of the supplementary characteristics that may be most useful in admissions or about how to assess them. It is quite possible that a primary need is to encourage more effective use of information already available by strengthening procedures for the subjective appraisal of applicants' personal qualities and accomplishments.

During the past few months, research scientists have made progress in organizing cooperative studies of personal qualities as factors in admissions, working with faculty and administrators in secondary schools, colleges, and graduate and professional schools throughout the country.

Inhibiting Factors

This examination of the admission process is but one of several ETS efforts to help make educational opportunities accessible to all who

Research psychologists K. Patricia Cross and Richard E. Peterson review lifelong learning survey, at ETS in Berkeley.

can benefit from them. Another is the array of projects aimed at overcoming the factors—such as unmet needs for financial aid and inadequate educational and career guidance— that inhibit access. Among these projects are the System of Interactive Guidance and Information, or SIGI (see page 10), and the College Board-sponsored Career Skills Assessment Program (see page 9).

Similarly, studies of student financial aid problems are continually undertaken by ETS staff. In 1977-78, a study sponsored by the Graduate and Professional Financial Aid Council (GAPFAC) examined the burden faced by graduate students when they start their careers at modest salaries and must repay education loans at fairly high monthly rates. The study resulted in the proposal of a new system, the Graduated Repayment Option (GRO), that would extend the repayment period from the usual 10 years to 15 and feature more reasonable monthly installments that increase as income does.

Educational Quality

Improving the quality of education is the prime purpose of many ETS research projects, from helping schools establish performance standards and evaluate programs to identifying and developing techniques for assessing outcomes of higher education.

For example, as part of the Basic Skills Assessment program (see page 7), ETS devised techniques for schools to use in setting local standards of student performance that are compatible with community attitudes and expectations. The testing program itself does more than measure each student against those standards. It also provides each school or school system with an aggregate indication of instructional strengths and weaknesses, thus pointing out needed improvements.

Luis Laosa, director of a longitudinal study of Mexican- American children, discusses the project with research assistant Judith Morgan.

In New York City, ETS research scientists continued to work with the school system to develop an accountability plan to hold the schools responsible for student learning and to correct instructional deficiencies by identifying strong programs and disseminating information abut them. For this purpose, student performance is assessed against a grade-level standard rather than compared with what peers have actually learned. The goal is to set proficiency levels for the entire school system.

Academic Competence

In 1977-78, a project conducted by ETS staff in Berkeley tested 1,150 college students in four areas of academic competence that are usually associated with general education. They are the analytic, communication, and synthesizing skills, and cultural awareness.

The 48-item test, developed with the help of faculty members at 15 California universities and colleges, called for brief, free-response answers. All questions were related to one of four content areas: mathematics and natural science, the humanities and fine arts, the social and behavioral sciences, and political science and history. Each of the cooperating institutions employs one of three major approaches to structuring the general education curriculum, and the study is expected to yield data that will be useful in evaluating the relative effectiveness of these approaches.

More Effective Education

Quality and effectiveness, while clearly intertwined, are not synonymous. As ETS uses the term, *effectiveness* applies to such aspects of the educational enterprise as productivity, the efficient use of resources, and the equitable distribution of them to both school systems and students.

One example is the application of technological advancements to administrative and instructional needs. What computers can and cannot do in the classroom is a major concern of a number of ETS research scientists.

CAI in Los Angeles

ETS researchers completed, during 1977-78, the second year of a four-year longitudinal study of computer-assisted instruction (CAI) in Los Angeles City Schools, with support from the National Institute of Education. Computer terminals are used by second, fourth, and sixth-grade students in four schools, to study reading and mathematics.

Participating students are being tested before and after each year's CAI experience, and ETS staff observe each fourth-grade classroom—there are about 25 of them—five times each year. Nonparticipating classes are monitored for purposes of comparison.

One of the most systematic and long-term studies of CAI ever undertaken, the project is intended to determine how valuable an instructional tool the computer can be—and why—and whether its use is economically feasible. The study will continue through 1980.

Program Evaluation

One way of assessing the effectiveness of education is systematic program evaluation, and ETS has been among the pioneers in this field. Its evaluations of *Sesame Street* and *The Electric Company*, as well as several programs of computer-assisted instruction, have been milestones in the development of these forms of educational technology.

ETS research scientists also design and conduct evaluations of programs in bilingual education, reading instruction, career education, and special education.

In 1977-78, for the Danforth Foundation, staff initiated a study that will synthesize information about effective techniques for evaluating programs in moral education in the schools.

School Financing

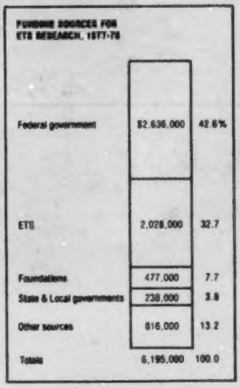
With grants from the Ford Foundation, among others, the ETS Education Policy Research Institute (EPRI) continued to build a program of research into state methods of financing public school systems. In 1977-78, a major study was conducted in New Jersey, where a court found local school financing practices unconstitutional and ordered redistribution of funds to provide a "thorough and efficient" education for all citizens.

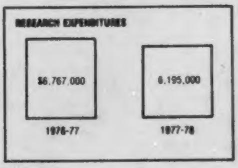
EPRI also examined the politics of school finance reform movements in various states, including Florida, Kansas, Maine, Michigan, and Minnesota. At the end of 1977-78, EPRI staff were analyzing the campaign for California Proposition 13 and the voting patterns that produced its resounding victory.

Human Development

A primary thrust of ETS research over the years has been the study of cognitive styles, the differing ways individuals organize and process information. While much of this work is basic research, practical applications to the teaching-and-learning process are constantly sought. For example, a community college study indicated that teachers are viewed more favorably by students whose cognitive styles are similar to their own.

Many other aspects of human learning and development are also being studied. They include close observation of infants and young children to discover what they learn and when, how family structure affects their development, the ways they acquire language skills, and the factors that influence personality and social behavior.





Learning Sex Roles

In 1977-78, a research team continued to examine sex-role stereotyping among youngsters in the New York City public schools under a grant from the National Institute of Mental Health. Their purpose is to determine whether, in light of the women's movement, traditional stereotypes are disappearing.

Results of another study seem to show that they are not. Based on results of a questionnaire administered to some 1,000 fourth and fifth-grade students in 40 central New Jersey classrooms, it was learned that a majority of the boys—and many of the girls—believed that, while intelligence is not sex-related, leadership is essentially the province of the male. The research scientist conducting the study is now working with teachers on instructional materials that will reduce the influence of stereotypes.

Special Populations

Identifying the educational needs of various population groups—racial minorities, adults, women, the economically disadvantaged, the culturally different, the gifted, the handicapped—claims the attention of many ETS research scientists. Several major projects of this sort were under way in 1977-78. Among them:

- A longitudinal study of the cognitive and social development of young Mexican Americans. Focusing on the effects of cultural and linguistic factors, the research team is studying preschool children in their homes and will track them through the early grades. A primary purpose is to learn whether—and if so, how—the development of bilingual children differs significantly from that of the monolingual majority.

- A nationwide survey of the continuing education needs and interests of adults and the programs, resources, and services available to them. Directed by research staff at the Berkeley office, the project will result in 1979 publication of the *Lifelong Learning Sourcebook* by Jossey-Bass, Inc., of San Francisco.
- A study to determine what factors encourage and discourage girls from studying advanced high school mathematics and science courses that are essential stepping stones to careers in the physical sciences and engineering. The critical factors were found to be the attitudes and behavior of parents, teachers, and guidance counselors. About 200 girls participated in the study, which was supported by the College Board and the National Science Foundation.

Exceptional Children

Under a five-year grant from USOE's Bureau of Education for the Handicapped, ETS established the Institute for the Study of Exceptional Children (ISEC), one of four institutes comprising the national Early Childhood Research Center. The others are located at the Universities of North Carolina, Kansas, and California at Los Angeles.

The new institute joins the research capabilities of the ETS Infant Laboratory with the medical expertise of the pediatrics department of Roosevelt Hospital in New York City. ISEC also works closely with the Coalition of Administrators of New Jersey Infant Programs, an umbrella organization for primary care programs that serve handicapped and at-risk children around the state. Thus, the institute crosses the usual boundaries between special education, pediatric medicine, and developmental psychology.

With four units—Detection; Research; Intervention and Evaluation; and Products, Delivery, and Training—ISEC focuses on early assessment and intervention for handicapped and at-risk children ranging in age from newborn to three years. Workshops and training sessions for professionals who work with the handicapped will be offered by ISEC at ETS.

Methodological Research

Underlying all ETS research, testing, and evaluation activities there is a solid foundation of methodological research. Over the past 25 years, for example, ETS research scientists have been at the forefront in the development and application of psychometric theory. Repeatedly, their work has contributed significantly to the refinement of educational measurement.

Currently, a team of research scientists and outside consultants is conducting a thorough examination of all the statistical practices ETS uses in developing and validating tests. The purpose is to assure that the organization's statistical methods are as advanced as the state of the art allows.

Michael Lewis, director, Institute for the Study of Exceptional Children, examines data with assistant Mary Katherine Hawryluk, center, and predoctoral fellow Laurie Waite.

Psychologist Jonathan Warren explains a current project to Berkeley colleagues (seated, from left) Michael Walsh, Ken Johnson, Judith Bonnett Hirabayashi.

Staff & Administration

Snowstorms closed ETS in Princeton twice during the winter of 1977-78, forcing staff to work overtime to meet critical deadlines for test registration, scoring, and reporting. The unprecedented closings came during a year marked by continued growth in many programs, increased workload and productivity, and extensive review and evaluation of activities across the organization.

Pat MacCrabie, compensation director, discusses procedures with assistant director Donald Bickel.

In the continuing effort to be accountable to its various constituencies, ETS gave considerable attention to controlling the quality of its products and the efficiency of its procedures. The major thrust of these endeavors was a systematic audit of all programs and activities, using the standards set forth in the *Principles, Policies, and Procedural Guidelines Regarding ETS Products and Services*, adopted the previous year.

When statistical errors were discovered in the score reports for a few test administrations, ETS moved swiftly to notify the affected students and institutions and to issue corrected scores. Procedures were tightened to reduce the chance of similar occurrences in the future. The organization also took steps to sharpen its test security measures following revelations that some test-takers had been coached with materials including actual test items.

Operations & Finance

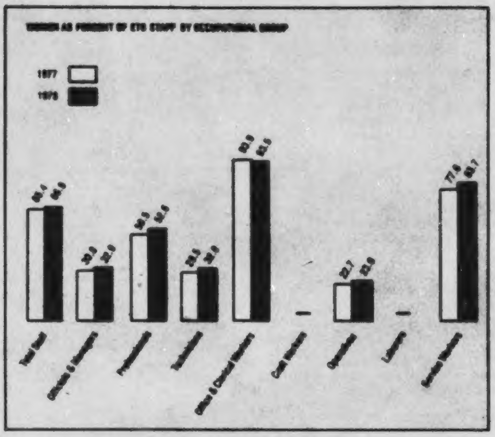
Construction of an extensive addition to Wood Hall on the main Princeton site was nearing completion at the close of 1977-78. It was designed to provide office space for about 350 staff, primarily engaged in student financial aid services, who previously occupied leased offices at nearby Research Park.

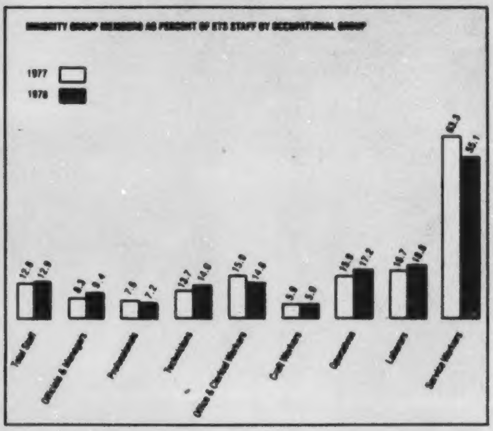
During the year, plans were drawn for the continued refinement and expansion of ETS' technological capabilities. New computers were to be installed later in 1978, and minicomputers were adapted to a variety of clerical and operational functions, including the processing of publication orders and inventories, and of general office work.

Dr. Anne Anastasi, a pioneer in the field of differential psychology, received the eighth annual ETS Award for Distinguished Service to Measurement at the 1977 Invitational Conference in New York. Now retired, she was on the Fordham University faculty.

Financially, 1977-78 was a satisfactory year, as indicated by the report of Coopers & Lybrand, certified public accountants, and documented in the supporting financial statements (see pages 11-14). Given the continuing inflation in the national economy, and the growth of several testing programs, total expenses were 13.3 percent above the 1976-77 level. Total revenue kept pace, rising 13.7 percent. The excess of revenue over expenditures increased from $901,022 in 1976-77 to $1,315,649 in 1977-78.

In its fifth year, the Henry Chauncey Conference Center continued to attract educational meetings and seminars from across





the nation and from other countries. The annual Bilderberg Meeting drew more than 100 international figures to the center for informal discussions of global trends and issues in April 1978.

Recruiting and Promoting

Equality remained a primary goal of staff recruitment and promotion practices in 1977-78. Under its Affirmative Action compliance programs, ETS continued to emphasize fairness in the hiring and advancement of employees, particularly women, minority group members, the handicapped, and Vietnam-era veterans. The organization exceeded national labor force participation rates for women and minorities in the officials-and-managers category, and for women in the professionals category.

During the year, supervisors participated in a series of workshops aimed at familiarizing them with the ETS Equal Employment Opportunity policy and Affirmative Action programs. Special attention was devoted to the legal backgrounds of the programs and guidelines for compliance.

Staff Development

Education and training programs for ETS grew more numerous and diversified in 1977-78. A great variety of seminars were offered, featuring guest speakers on issues and trends in education, measurement, and human development. Data processing staff in Princeton presented a wide range of courses, some developed inhouse and some purchased in package form, to employees across the organization.

At the same time, three postsecondary institutions—Mercer County Community College, Trenton State College, and Rutgers University—offered credit courses to staff at the Rosedale Road site in Princeton.

Some Current ETS Publications

Periodicals

CATC Digest, a quarterly newsletter for persons interested in computer-assisted test construction.

ETS Developments, a quarterly newsletter about ETS measurement, research, and related activities of general interest.

Findings, a semiannual report of ETS research in postsecondary education.

International Newsletter, a vehicle for the exchange of measurement information from various countries, published semiannually.

Selected Reports

A Guide to ETS Activities Related to Minority and Disadvantaged Groups, 1970-77, a compendium prepared by Audre Griffiths and John A. Winterbottom and published by the ETS Office for Minority Education.

Children of Poverty: Overcoming the Barriers, an eight-page summary of a report by Virginia C. Shipman on the social and educational factors that are characteristic of low- and high-achieving disadvantaged children, part of the ETS-Head Start Longitudinal Study.

**College Achievement Through Self-Help* by Luis Nieves, a manual designed to help minority students cope with the special problems they face in adapting to campus life.

Coping on Campus: The Problems of Minority College Students, an eight-page summary of the results of a study conducted by the ETS Office for Minority Education for the Graduate Record Examinations Board.

**Educational Measurement and the Law*, proceedings of the 1977 ETS Invitational Conference.

Focus: Learning to Read, fourth in a series of occasional papers about current issues and problems in education and measurement and related work at ETS.

Focus: The Concern for Writing, fifth in the series.

**How To Get College Credit for What You Have Learned as a Homemaker and Volunteer* by Ruth B. Ekstrom, Abigail M. Harris, and Marlaine E. Lockheed, a guide for women who plan to enter or return to college and seek credit for prior learning.

Money and Education: Where Did the 400 Million Dollars Go? by Margaret Goertz, report of a study of school finance reform in New Jersey, conducted by the ETS Education Policy Research Institute.

Public Information Brochures

About ETS, a brief overview of ETS activities in testing, measurement, related services, and educational research.

Assessing Occupational Competence, a description of the programs and procedures of the ETS Center for Occupational and Professional Assessment (COPA).

ETS in Fact, 1979, a collection of facts and figures about ETS testing programs, services, research, finances, and regional offices.

ETS International Activities, a brief survey of ETS programs and projects involving agencies, institutions, and individuals of other nations.

ETS Publications, 1979, a catalog of current publications including research reports and testing program materials, with form and instructions for ordering.

Keeping Up, a folder containing a catalog of ETS Programs of Continuing Education and a booklet detailing available instructional services.

---

*There is a charge for each publication marked with an asterisk. To order publications or obtain additional information about them, please contact the Information Division, ETS, Princeton, NJ 08541; 609-921-9000.

A Partial List of ETS Clients, 1977–78

ALPHA Center for Health Planning
Albright College, Pennsylvania
American Assembly for Collegiate Schools of Business
American Association of Colleges of Pharmacy
American Association of Colleges of Podiatric Medicine
American Association of Teachers of French
American Association of Teachers of German
American Bankers Association
American Board of Obstetrics and Gynecology
American Council on Education
American Federation of Teachers
American Foundation for Negro Affairs
American Institute of Planners
American Institutes for Research
American Journal of Nursing Company
American Nurses' Association
American Pharmaceutical Association
American Production and Inventory Control Society
American Psychiatric Association
American Registry of Radiologic Technologists
American Speech and Hearing Foundation
Applied Data Research, Incorporated
Arabian-American Oil Company Education Department
Asia Foundation
Association of American Medical Colleges
Association of Collegiate Schools of Architecture
Atlanta Area Association of Independent Schools, Georgia

Bank of Canada
Bermuda Department of Education
Bermuda Finance Department
Bilingual Education Services Center
Biomedia, Incorporated
Borough of Manhattan Community College, New York
Boston Public Schools, Massachusetts
Brookline Public Schools, Massachusetts

CASL, Incorporated
California State Universities and Colleges
Caribbean Examinations Council
Carnegie Corporation of New York
Carnegie Council on Policy Studies in Higher Education
Carnegie-Mellon University
Charles County Community College, Maryland
Chicago Educational Television Association (WTTW)
City of Atlanta
City of Chicago
City of New York
City of Philadelphia
College Entrance Examination Board
College of Charleston, South Carolina
College of Medicine and Dentistry of New Jersey
Commission on Graduates of Foreign Nursing Schools
Committee of Bar Examiners, State of California
Consumers Union
Council of Graduate Schools in the United States
Council on Legal Education for Professional Responsibility, Incorporated
Creighton University, Nebraska
Danforth Foundation
Departments or agencies of the following states or commonwealths:
Alaska
California
Colorado
Connecticut
Delaware
District of Columbia
Florida
Georgia
Hawaii
Illinois
Indiana
Iowa
Kansas
Kentucky
Louisiana
Maine
Maryland
Massachusetts
Michigan
Missouri
Montana
Nebraska
Nevada
New Hampshire
New Jersey
New York
North Carolina
North Dakota
Oregon
Pennsylvania
South Carolina
South Dakota
Tennessee
Vermont
Virginia
Virgin Islands
Wisconsin
Wyoming

De Pauw University, Indiana
Development and Resources Corporation

Eastman Kodak Company
Educational Commission for Foreign Medical Graduates
Educational Records Bureau
EXXON Foundation

Fairfax School District, California
Fayetteville State University, North Carolina
Flint Community Schools, Michigan
Florida Junior College at Jacksonville
Ford Foundation
Freehold Regional High School District, New Jersey

Garvey School District, California
Government of Trinidad and Tobago
Graduate Management Admission Council
Graduate Record Examinations Board
Grant Foundation

High/Scope Educational Research Foundation
Hospital Financial Management Association

Instituto Tecnológico y de Estudios Superiores de Monterrey
International Business Machines Corporation
International Transactional Analysis Association

Japan-United States Friendship Commission
Jersey City Public Schools, New Jersey
Johns Hopkins University
Josiah Macy Foundation

Kern Union High School District, Kern County, California

- Law School Admission Council
- Lilly Endowment
- Linfield College, Oregon
- Livingstone College, North Carolina
- Lompoc Unified School District, California
- Los Angeles County, California
- Los Angeles Unified School District

- Mercer County Community College, New Jersey

- National Association of Boards of Pharmacy
- National Association of Housing and Redevelopment Officials
- National Association of Independent Schools
- National Association of Secondary School Principals
- National Association of Social Workers
- National Board for Respiratory Therapy
- National Board of Podiatry Examiners
- National Business Forms Association
- National Catholic Education Association
- National Conference of Bar Examiners
- National Council of Architectural Registration Boards
- National Council of Engineering Examiners
- National Furniture Warehouseman's Association
- National Institute for Automotive Service Excellence
- National Merit Scholarship Corporation
- Newark School District, Delaware
- Newark Board of Education, New Jersey
- New Jersey School Boards Association
- New York Stock Exchange
- Northeastern University, Massachusetts
- Norwalk Board of Education, Connecticut

- Opticians Association of America
- Oxnard Union High School District, California

- PACE Institute, Incorporated
- Pasadena Unified School District, California
- Pompton Lakes Board of Education, New Jersey
- Princeton University, New Jersey
- Professional Golfers' Association of America
- Province of New Brunswick, Canada
- Province of Nova Scotia, Canada

- Radio Corporation of America
- Robert Wood Johnson Foundation
- Rockefeller Family Fund
- Rutgers, The State University, New Jersey

- Santa Monica Unified School District, California
- School District of Philadelphia, Pennsylvania
- Scientific American
- Secondary School Admission Test Board
- Secondary School Research Program
- Society of Actuaries
- Southeastern Tidewater Area Manpower Authority

- Texas Education Agency
- Trenton State College, New Jersey

- Union College, New Jersey
- **United States departments and agencies:**
  - ACTION
    - Peace Corps
  - Civil Service Commission
  - Council on Environmental Quality
  - Department of Defense
    - Department of the Army
      - Army Research Institute for the Behavioral and Social Sciences
      - Army ROTC
      - Defense Advance Research Projects Agency
    - Department of the Air Force
      - Air Force ROTC
    - Department of the Navy
      - Defense Activity for Nontraditional Education Support
      - Marine Corps
    - Office of General Counsel
    - National Security Agency
  - Department of Health, Education and Welfare
    - Office of the Assistant Secretary for Education
      - Fund for the Improvement of Postsecondary Education
      - Office of Education
      - National Institute of Education
    - Office of Human Development
    - Public Health Service
      - Health Resources Administration
      - National Institutes of Health
  - Department of Labor
    - Employment and Training Administration
  - Department of State
    - Agency for International Development
    - Board of Examiners of the Foreign Service
  - Department of Transportation
    - Coast Guard
  - Environmental Protection Agency
  - International Communication Agency
  - National Endowment for the Humanities
  - National Science Foundation
  - National Security Council
    - Central Intelligence Agency
- United States District Court for the Northern District of Texas, Dallas Division
- University of California Center for Research and Development in Higher Education
- University of London
- University of Maryland
- University of Southern California, School of Medicine
- Uniworld Group, Incorporated

- Whitman College, Washington
- World University, Puerto Rico

ETS Trustees, 1977–78

Board of Trustees

Albert N. Whiting, *Chairman*
Chancellor, North Carolina Central University

Elie Abel
Dean, Graduate School of Journalism, Columbia University

M. Anne Campbell
Commissioner of Education, State of Nebraska

Thelma T. Daley
Career Education Specialist, Board of Education of Baltimore County

John W. Hennessey, Jr.
Professor of Management, The Amos Tuck School of Business Administration, Dartmouth College

Wayne H. Holtzman
President, Hogg Foundation for Mental Health, and Professor of Psychology and Education, The University of Texas at Austin

Clark Kerr
Chairman and Staff Director, Carnegie Council on Policy Studies in Higher Education

C. Peter Magrath
President, University of Minnesota

Sidney P. Marland, Jr.
President, The College Board

Peter Masiko, Jr.
President, Miami-Dade Community College

Gertrude G. Michelson
Senior Vice President, Macy's New York

Charles D. O'Connell, Jr.
Vice President and Dean of Students, The University of Chicago

Jack W. Peltason
President, American Council on Education

Virginia B. Smith
President, Vassar College

Stephen H. Spurr
Professor, Lyndon B. Johnson School of Public Affairs, The University of Texas at Austin

William W. Turnbull
President, Educational Testing Service

Executive Committee

Stephen H. Spurr, *Chairman*

Elie Abel

John W. Hennessey, Jr.

Wayne H. Holtzman

Charles D. O'Connell, Jr.

William W. Turnbull

Albert N. Whiting

Finance Committee

John W. Hennessey, Jr., *Chairman*

Kurt M. Hertzfeld
Treasurer, Amherst College

Peter Masiko, Jr.

Gertrude G. Michelson

Stephen H. Spurr

William W. Turnbull

Audit Committee

John W. Hennessey, Jr., *Chairman*

Gertrude G. Michelson

Stephen H. Spurr

Committee on Research and Development

Wayne H. Holtzman, *Chairman*

Thelma T. Daley

Robert Glaser
Co-director, Learning Research and Development Center, University of Pittsburgh

Philip W. Jackson
Chairman, Department of Education, The University of Chicago

Christine McGuire
Associate Director for Research and Development, Center for Educational Development, College of Medicine, University of Illinois

Wilbert J. McKeachie
Professor of Psychology, University of Michigan

Sister Joel Read
President, Alverno College

Richard E. Schutz
Executive Director, Southwest Regional Laboratory for Educational Research and Development

Virginia B. Smith

William W. Turnbull

ETS Officers, 1977–78

President

William W. Turnbull

Executive Vice President

Robert J. Solomon

Senior Vice Presidents

Scarvia B. Anderson
David J. Brodsky
Winton H. Manning
E. Belvin Williams

Vice Presidents

Richard L. Burns
Jack R. Childress
Marion G. Epstein
John S. Helmick
Samuel J. Messick
Robert E. Smith
Bernard L. Tchorni

Treasurer

Donald S. Grant

General Counsel

John S. Kramer

Secretary

Jane D. Wirsig

Assistant Vice Presidents

L. Lynnwood Aris
Antoinette M. Orsi

Assistant General Counsels

Russell W. Martin, Jr.
Stanford H. von Mayrhauser

Assistant Treasurer and Controller

Frank R. Henderson

Regional Offices

Midwestern Region

1 American Plaza
Evanston, IL 60201

Northeastern Region

Mid-Atlantic Office
Rosedale Road
Princeton, NJ 08541

New England Office
2 Sun Life Executive Park
100 Worcester Road
Wellesley Hills, MA 02181

Southern Region

Suite 1040
3445 Peachtree Road, NE
Atlanta, GA 30326

Southwestern Region

Suite 100, Southwest Tower
211 East Seventh Street
Austin, TX 78701

Western Region

Berkeley Office
1947 Center Street
Berkeley, CA 94704

Los Angeles Office
Room 216, 2200 Merton Avenue
Los Angeles, CA 90041

Puerto Rico

Suite 1115, Banco Popular
209 Muñoz Rivera Avenue
Hato Rey, PR 00918

Washington Office

Suite 300
1800 Massachusetts Avenue, NW
Washington, DC 20036

EDUCATIONAL TESTING SERVICE

Finances...Where the Money Comes From and Where It Goes

Educational Testing Service is a nonprofit organization that provides testing, financial aid and transcript services to students and institutions; conducts research and development in educational measurement; and provides advisory and instructional services in the proper uses of measurement. In the fiscal year ending June 30, 1979, ETS earned income of $94 million on expenses of $93 million, resulting in net revenues of just over $1 million--or 1.2% of expenses.

Income

Most of ETS' income--close to $85 million, or 90%--came from the services that it provides in conducting program services for various test sponsors, e.g. the College Board. These services include developing tests, printing and distributing the tests, administering them to over 7 million individuals annually throughout the world, scoring the tests and reporting those scores to test takers and institutions. To ensure test quality, ETS also conducts statistical analyses and research on test results and publishes extensive information about the proper uses of tests and test scores.

Most testing programs are conducted by ETS under contract with client groups who make policy decisions: what skills and knowledge will be measured, how often tests will be offered and how much they will cost the test taker. Usually, the client reimburses ETS for its costs and pays a service charge out of test fees, keeping the remainder for its own activities.

In 1978-79, about 53% of program income came from College Board programs; approximately 19% came from graduate and professional school programs, e.g. the Graduate Record Examination and the Law School Admissions Test; another 13% came from the testing of teachers and other professionals and the remaining 15% from other testing activities at various educational levels (Attachment A). Of this last amount, approximately $500,000 or well under 1% of total program income was received for the sale of published tests, e.g. Cooperative Tests.

Under its Charter and By-Laws, in addition to providing services for testing programs, ETS is mandated to conduct research in measurement, develop new services...."in the field of tests and measurement", "promote the understanding of...educational measurement". "counsel test users on the techniques of measurement", "undertake...advisory services" and "serve...as a clearinghouse of information about research". Thus, in addition to specific research conducted for testing programs, ETS conducts a broad program of research and development in measurement and provides advisory, consulting and instructional services.

In 1978-79 nearly $8 million, or just over 8% of ETS' total income, was received from outside sources in support of these activities. Approximately three-quarters of that amount came from projects conducted for government--state, federal and local agencies, about 16% from conducting specific program research and providing advice to schools and colleges and the remainder from studies and activities funded by foundations. More than 10,000 research reports were distributed last year.

Nearly $1½ million in income, slightly over 1½% of the total, came from other sources--primarily from fees charged to users of the Henry Chauncey Conference Center and from returns on investment in marketable securities. (See Audited Statement of Revenue and Expenditures, Attachment B.)

Expenses

In the fiscal year ended last June, ETS' total expenses amounted to $93 million. Not surprisingly, the great majority of these expenses, 83%, were devoted to providing programs and services for more than 300 separate clients. Over 15% was spent for programs of research, development and advisory and instructional activities. Less than 2% of expenses were devoted to operating the Conference Center.

On a functional basis, close to $49 million or 53% of ETS' expenses were devoted to salaries, wages and fringe benefits. Of this amount $40 million was for the salaries and wages of the 2300+ permanent staff members and for the payroll of the temporary staff which exceeded 1100 people at peak time--equivalent in hours worked to about 630 full-time employees. Average salary for the total staff was approximately $13,600 compared to about $13,800 derived from the Bureau of the Census' Statistical Abstract which reports income of all wage earners nationwide.

ETS salaries are comparable to salaries established for jobs requiring skills and talents similar to those needed in ETS' work. For example, the salaries of ETS professional research staff are compared with those of people conducting similar work in universities and research institutions; salaries of data processing staff, program managers, administrators, secretaries and clerical staff are compared with those of their counterparts in other institutions and organizations. (See Attachment C.) The ETS President's salary of $88,000 is comparable to the salaries of chief executive officers of major educational institutions, nonprofit organizations (including research institutes and foundations) and other service organizations of similar size.

The rest of ETS' major expenses, in order of size are: about $10 million to pay some 38,000 test administrators, essay examination readers, question writers and other professional consultants who help in ETS' work; ETS' printing bill—nearly $9 million—for test booklets, answer sheets and published interpretative materials and reports sent out last year; about $7 million spent on the freight and postage for sending and receiving those materials; $4 million paid for computer rental and services; $2 million for depreciation on buildings and owned equipment; $1½ million for the expenses of utilities and property taxes and a telephone bill of about $1½ million, much of which arises from informing and responding to inquiries from students and institutions. These represent 90% of ETS' costs. The remaining operating and administrative expenses--the costs of books, supplies, professional travel and the like, are enumerated in hundreds of individual accounts too extensive to list here.

Net Revenues

ETS net revenues, i.e. the excess of income over expenses, amounted to $1.1 million in 1978-79--1.2% of expenses.

In order to help support its other charter-mandated activities and to provide funds for the plant and equipment needed to carry on its work, ETS charges a fee beyond costs to its program clients as well as for some of its contract research. Fees are negotiated independently with each of its clients and in each of its contracts, but they have consistently averaged between 9 and 10% of program expenses in the last ten years. Of those program fees, which amounted in total to $7¼ million in 1978-79, nearly 85%, or about $6.3 million (in addition to over $7.8 million in outside funds) was used to support research, development and advisory and instructional services last year. The other 15% of program fees--i.e. the $1.1 million in net revenues--was used for additions to operating buildings, capital equipment and working capital.

Net Assets

At the end of fiscal year 1978-79 ETS had net assets, or a Fund Balance, of nearly $28 million. This represents the total amount of capital contributed to ETS at its founding (about $1¼ million) plus the accumulated net revenues from 31 years of operation. (See Audited Balance Sheet, Attachment D.)
Of this amount, about $26¼ million, or 95% was put back into property, plant, furniture and equipment. Most of this "reinvestment" occurred after the mid-50s when the surge of post-war students sought admission to institutions of higher education and ETS was pressed to serve them with an antiquated plant scattered in more than a half-dozen locations in the downtown Princeton area.

In 1955 ETS bought 365 acres of farmland (at approximately $700 an acre) just outside Princeton and in 1958 completed construction of the first of a set of buildings that were the base of its present modern and efficient plant. The ETS plant now comprises some ten buildings, used for offices, research, clerical and computer operations and shipping and warehousing. All buildings were constructed after being competitively bid and ETS' annual occupancy rate at $7 per square foot--including depreciation, utilities, maintenance and the real estate taxes that it pays--compares very favorably with office and operating facilities costs in the region.

The Henry Chauncey Conference Center, first opened in August of 1973, was built to house the hundreds of people who come to Princeton to work with ETS annually on test development committees and program councils, as well as educators and consultants who advise ETS on its programs of measurement and research. The Center was also created to serve as a base for ETS' programs of instruction and to house the many foreign visitors and students who come to ETS to learn about measurement theory and practice. It is a popular site for other educational groups and provides a place for the exchange of problems and ideas for the improvement of the educational process.

From the beginning of the project it was recognized that in order to operate such a facility without imposing a financial burden on the rest of ETS' activities, it would be necessary to accommodate outside groups until ETS' own growing needs reached a level to sustain the Center economically. Thus, it was concluded that a multi-purpose educational center with facilities adequate to attract many groups, including attractive recreational facilities typically found at conference centers, would need to be constructed. The 100-room Center, built at a cost of three million dollars, has operated at an annual deficit since its opening. The deficit in 1978-79 was about $150,000--well under 2/10 of 1% of total income. This deficit has been reduced annually to the point where the Center is projected to be totally self-supporting within the next 24 months. It should be noted that the primary purpose of the Center is and always has been to facilitate ETS activities in testing, research and instruction, that it was built out of earned resources from ETS programs and services, and that the operating deficit reflects a full allocation of costs, including taxes and depreciation.

Apart from its plant and equipment and the working capital (amounts owed to vendors and others, less amounts coming in from fees and contracts or amounts held for providing future services) required by every organization, ETS has no financial assets other than its portfolio of marketable securities, valued at cost on June 30, 1979 at about $4.8 million. This amount, which is held for equipment replacement and operating or financial contingencies, represents less than three weeks of operating expense at ETS' current level of activity. As noted above, the return on those investments helps to support all of ETS' activities. In 1978-79 the return was about $300,000--or approximately double the amount required to support the Conference Center.

Costs and Efficiency

ETS management is continually challenged by the problems of keeping costs (ultimately reflected in student fees) as low as possible while providing extensive and complex services for millions of students and thousands of institutions, a challenge compounded by the critical requirement that the processed information be as accurate and timely as is humanly possible. Since all of ETS' services are labor-intensive, requiring the close attention of people for their effective accomplishment, the majority of ETS' expenses are payroll and benefits costs, as was mentioned earlier. Constant attention to maintaining efficient work facilities and to opportunities for automating those activities that can be more accurately and swiftly accomplished by computers and other labor-saving devices has paid off. As a measure of that success, in four of the largest programs that ETS conducts for its clients--the Admissions Testing Program of the College Board, the Graduate Record Exam, the Law School Admission Test and the Graduate Management Admission Test--student test fees, which reflect costs, have risen on the average less than 50% in the last 25 years while the Cost of Living Index has just about tripled during that time (see Attachment E). In the last decade, while the cost of living nearly doubled and the number of students taking the College Board's Scholastic Aptitude Test (a bellwether program) dropped by about 6%, the fee for taking the SAT rose by only 45%.

As has been pointed out earlier, some 90% of ETS' income comes from programs and services and the majority of that income comes indirectly from student fees. Attachment F is an illustration of where the ETS income dollar comes from and where that dollar is spent. Attachment G shows where ETS spends its portion of "test-taker" fees, based on the expenses of four major national testing programs. Most of the fee is spent on supervisors' honoraria, processing of tests and on the production, shipping and mailing of test booklets, answer sheets, test information and reports. At present about 6 cents on the dollar go to develop and analyze tests. Depending on the nature of the particular test--aptitude or achievement, broad or highly-specialized subject matter--the cost of test development of a single form of a test can range from $30,000 to $100,000.

Another measure of ETS' operational efficiency is its growth in staff compared with the work required to be accomplished. In the last five years, ETS staff has grown 19% while its workload, as measured by processing volumes in its major programs, has increased at almost double that rate.

ETS "overhead" or indirect costs have remained in a band between 15% and 20% of total costs for the last ten years and have dropped steadily in the last five years from 19% to 16½%.

Finally, in the test of the marketplace, ETS has won a substantial portion of the contracts for which it has bid on a competitively priced basis.

Accountability

Since its inception ETS' finances have been audited annually by an independent external firm of accountants and auditors (presently Coopers & Lybrand) who report directly to the Audit Committee of the ETS Board of Trustees. ETS publishes an annual report which includes its audited financial statements. ETS financial performance is reviewed at least semi-annually by the Finance Committee--members of the Board with particular knowledge of the finances and management of major institutions. In addition, several of ETS' major clients, e.g. the College Board, conduct independent annual audits of ETS' financial management of their programs, as does the federal government.

Recommendations concerning the allocation of resources to the support of research and development are made by the Finance Committee to the full Board, based on research plans reviewed independently by the Research and Development Committee. The Finance Committee also reviews and approves all allocations of resources for the acquisition of plant and equipment and makes all final investment decisions which are reviewed and confirmed by the full Board. Summaries of ETS' contracts with its major clients have been available for inspection for the past five years.

FIVE-YEAR SUMMARY OF REVENUE

	1979	1978	1977	1976	1975
College Board Testing Programs and Services	$45,545,032	$38,170,392	$33,168,875	$30,809,832	$27,577,162
Graduate and Professional School Testing Programs	15,670,744	14,399,328	13,652,337	12,565,157	11,731,892
Testing of Teachers and Other Professionals	11,210,159	8,605,727	6,577,947	5,807,559	5,087,719
Other Testing Activities at Various Academic Levels	12,492,436	10,333,443	8,261,160	6,688,391	6,123,785
Total Testing Activities	84,918,371	71,508,890	61,660,319	55,870,939	50,520,558
Research, Development, and Instructional Services	7,837,513	6,860,671	7,202,403	5,819,161	7,003,362
Other Income	1,447,187	1,256,782	1,144,124	1,234,250	1,002,308
Total Revenue	$94,203,071	$79,626,343	$70,006,846	$62,954,350	$58,526,228

DRAFT

For discussion purposes only -
Subject to Revision

ATTACHMENT A

ATTACHMENT B

STATEMENTS OF REVENUE, EXPENDITURES AND FUND BALANCE
for the years ended June 30, 1979 and 1978

	1979	1978
Revenue:		
Testing activities	$84,918,371	$71,508,890
Research, development, advisory and instructional services (Note 5)	7,837,513	6,860,671
Other revenue	1,447,187	1,256,782
Total revenue	94,203,071	79,626,343
Expenditures:		
Salaries, wages and employee benefits	48,854,458	42,317,158
Test administrators, readers, item writers and other professional services	10,380,543	8,282,645
Printing and materials	8,123,[illegible]	7,272,207
Depreciation	1,932,609	1,692,371
Electronic equipment rental and services	4,066,[illegible]	3,321,664
Utilities and property taxes	[illegible]	[illegible]
Freight and postage	[illegible]	[illegible]
Telephone and telegraph	1,398,425	1,218,232
Other operating and administrative expenses (Notes 6 and 8)	9,700,[illegible]	6,480,[illegible]
Total expenditures	93,087,237	78,310,694
Excess of revenue over expenditures	1,115,834	1,315,649
Fund balance, beginning of year	26,771,717	25,456,068
Fund balance, end of year	$27,887,551	$26,771,717

See notes to financial statements.

DRAFT
FOR DISCUSSION PURPOSES ONLY —

ATTACHMENT C

EDUCATIONAL TESTING SERVICE

Representative Salary Comparisons 1978-79

ETS Salary Ranges		Outside Salary Ranges and Average Salaries	
Professional and Administrative Staff			
$26,360 - $43,440 Senior Research Scientist, Sr. Program Administrator, Sr. Division Director	A	Professor Assoc. Professor Asst. Professor	$34,450 average $25,950 average $20,890 average
$23,220 - $38,010 Research Scientist, Head Statistical Systems, Division Directors, Senior Test Examiners	B	U.S. Civil Service (GS 13) Scientists, Professionals, Senior Administrators	$27,433-$35,688
$20,520 - $33,560 Senior Research Associate, Test Examiner, Program Administrator	C	University Deans University Directors	$31,935 average $22,206 average
Data Processing Staff			
$20,620 - $33,540 Senior Systems Analyst, Sr. Program Analyst	D B	National Survey US Civil Service(GS 11)	$26,250 average $19,263-$25,044
$18,490 - $29,880 Staff Systems Analyst, Staff Program Analyst	D B	National Survey US Civil Service(GS 10)	$25,688 average $17,532-$22,788
Clerical Staff			
$8,920 - $13,360 Senior Clerical Assistant	E F B	Regional Clerical Salary National Clerical Salary US Civil Service (GS 4)	$10,226 average $10,816 average $9,391-$12,208
Total Staff			
$13,600 Average Salary	G	National Average - all full-time wage earners	$13,800

A American Ass'n. University Professors 1978-79 Adjusted to 11 months

B US Civil Service Salary Grades thru Aug. 1979

C College and University Personnel Association March 1979

D Info systems Magazine - June 1979

E Local Area Personnel Survey 1979

F Bureau of Labor Statistics

G US Bureau of Census - Statistical Abstract 1978

EDUCATIONAL TESTING SERVICE
BALANCE SHEETS
June 30, 1979 and 1978

ASSETS	1979	1978
Current assets:		
Cash	$ 1,524,833	$ 2,561,572
Short-term investments (approximates market)	6,403,000	5,600,000
Accounts receivable, less allowance for doubtful accounts: ($250,000 in 1979 and $100,000 in 1978)	4,335,221	3,464,821
Unbilled amounts under contracts	4,019,982	3,[illegible]
Prepaid expenses	1,681,974	1,408,162
Total current assets	17,964,410	16,[illegible]
Investments in marketable securities (market value 1979, $4,600,000 including in 1979, $730,000 as equipment replacement fund)	4,826,054	4,492,010
Property, plant, furniture and equipment, net (Note 2)	26,470,388	24,531,073
	$49,260,852	$45,385,666

LIABILITIES	1979	1978
Current liabilities:		
Accounts payable	$ 3,594,988	$ 3,205,421
Accrued expenses and other liabilities	5,052,061	3,729,759
Unexpended receipts:		
Under contracts	7,030,334	5,912,917
Under grants including advance payments	526,117	1,282,508
Total current liabilities	16,203,500	14,130,605
Deferred income (Note 3)	598,987	673,861
Unexpended receipts (Note 4)	4,570,814	3,809,483
Fund balance (Note 5)	27,887,551	26,771,717
	$49,260,852	$45,385,666

See notes to financial statements.

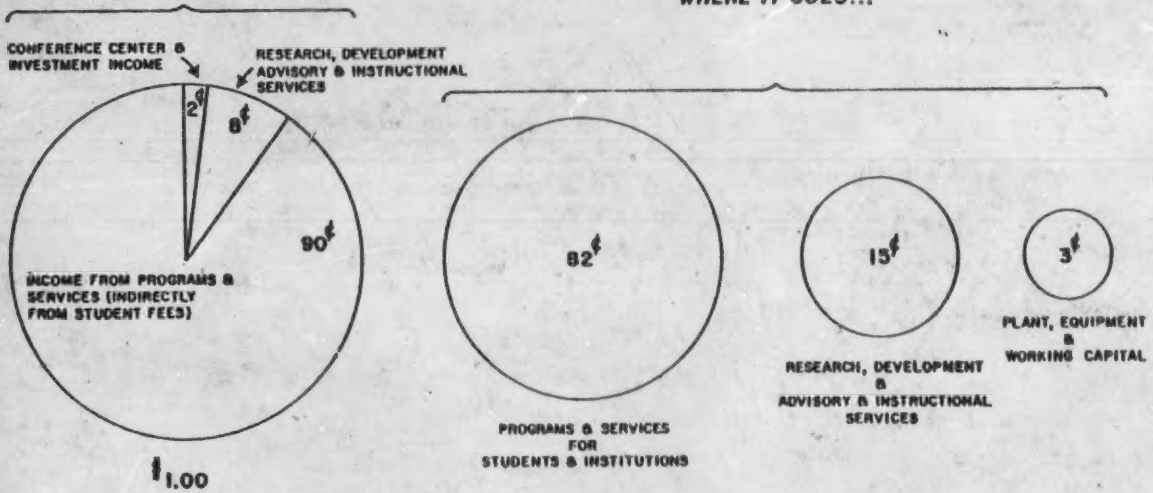
DRAFT
FOR DISCUSSION PURPOSES ONLY -
SUBJECT TO REVISION

ATTACHMENT D

ATTACHMENT E

CANDIDATE FEES IN FOUR MAJOR PROGRAMS vs. INFLATION

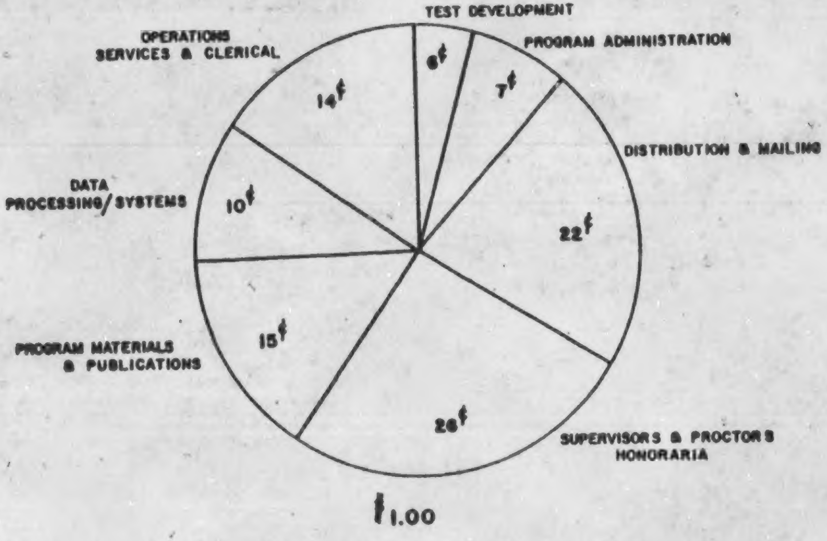
	1954-55 Fee	25 Years' Inflation @ 191.68%	1979-80 Actual Fee
Admissions Testing Program	$ 6.00	$17.50	$ 8.25
Graduate Management Admission Test	10.00	29.15	12.50
Graduate Record Examination	8.00	23.33	14.00
Law School Admission Test	10.00	29.15	15.00


WHERE THE ETS INCOME DOLLAR COMES FROM...

WHERE IT GOES...

ATTACHMENT G

WHERE THE TEST TAKER DOLLAR IS SPENT IN FOUR MAJOR PROGRAMS *



* COLLEGE BOARD ADMISSIONS TESTING PROGRAM
GRADUATE MANAGEMENT ADMISSIONS TEST
GRADUATE RECORD EXAMINATIONS
LAW SCHOOL ADMISSION TEST

PROGRAM RESEARCH PROGRESS REPORT

AUGUST 1979

THE COLLEGE BOARD
GRADUATE MANAGEMENT ADMISSION COUNCIL
GRADUATE RECORD EXAMINATIONS BOARD
LAW SCHOOL ADMISSION COUNCIL
TEST OF ENGLISH AS A FOREIGN LANGUAGE

PROGRAM RESEARCH PROGRESS REPORT

The Program Research Progress Report serves as a summary of program research projects of five major programs at Educational Testing Service: The College Board, Graduate Management Admission Council, Graduate Record Examinations Board, Law School Admission Council, and Test of English as a Foreign Language. This report recognizes that ETS research program sponsors often are interested in the research activities of other ETS programs. The report includes one-page summaries of projects in progress or projects recently completed and a listing of projects completed during the past five years. The summaries describe the purpose, progress to date, planned completion schedule, and findings in the case of completed projects.

For completed projects, citations are included so that those interested may obtain copies of reports where available. Requests for reports should be addressed to the individual program offices at ETS.

This report was prepared by the

Office of Program Research
Educational Testing Service
Princeton, New Jersey 08541

Questions concerning specific projects should be directed to the project directors or the respective program research coordinator at ETS:

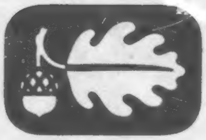
Hunter Breland --The College Board
Thomas Barrows --Graduate Management Admission Council
Mary Jo Clark --Graduate Record Examinations Board
Franklin Evans --Law School Admission Council
Donald Alderman--Test of English as a Foreign Language

Linda Johnson
Staff Associate

Warren W. Willingham
Assistant Vice President
for Program Research

A. PROGRESS REPORTS FOR RESEARCH PROJECTS OF

THE COLLEGE BOARD



Title: SAT Speededness Study

Project/Job Number: 142/15 Project Director: Franklin R. Evans

Purpose

The study will seek to determine (1) if the SAT is more speeded for minority group (Mexican-American and black) than white candidates, (2) if the SAT is more speeded for rural than urban candidates, and (3) if removing speededness from the SAT results in more valid or less valid scores for minority group candidates.

Progress (as of July, 1979)

Special tests were administered in 12 schools in January and February, 1972. Scores are on tape and data analysis is complete. Preliminary results indicate that removing speededness does not provide any differential benefit for any of the various groups. Various single administration indices of test speededness have been studied.

Schedule for Completion

A draft report has been prepared and submitted for review.

Title: Test of the Adequacy of Score Equating Models

Project/Job Number: 142/25 **Project Directors:** Gary L. Marco
E. Elizabeth Stewart

Purpose

To test the adequacy of standard ETS equating models and a variety of new models when sample and test characteristics are systematically varied.

Progress (as of August, 1979)

Scoring was completed in April 1977, and all equatings have been performed for each of the 466 equating situations. A total of 40 linear, two equipercentile (direct and frequency estimation), and two item characteristic curve (1- and 3-parameter logistic) equating models were used. Summary tables of equating results have been prepared. Comparative analysis of the curvilinear models have been completed and a report on this part of the study has been prepared. Comparative analysis of the linear models are in progress.

Schedule for Completion

The final reports are scheduled to be completed by October 1979. Final summary tables and graphs are in the process of being completed. A draft of the final report is in process except for the section on the interpretation of the comparative analysis of the linear models.

Findings

When total tests differ in difficulty, a curvilinear equating model is superior to a linear model. Because of the criterion problem, a choice cannot easily be made among the curvilinear models, except that the 1-parameter model tends to be superior when a test is equated to itself and it tends to be inferior when a test is equated to a different test. All models are satisfactory when conditions are optimum.

Title: Literature Review of Relationships Among Tests, Grades, and Life Success

Project/Job Number: 142/27 Project Director: Leonard Baird

Purpose

This review is designed to critically assess and synthesize the available research which bears on the question, "Do academic ability test scores and grades predict anything but grades; i.e., are they related to real life accomplishment?"

Progress (as of July, 1979)

Research from many sources has been examined for their relevance to the question above. Fifteen journals were systematically surveyed by examining the articles published in them during the last ten years. These articles often referred to other articles in journals outside the fifteen. In addition, relevant categories in Psychological Abstracts for the last five years were systematically reviewed. Finally, several leading figures in creativity research were asked to send any published or unpublished work that could help answer the question. These sources led to many reports which had to be examined carefully, because the relevant analyses were often hidden in appendices, and the question of concern in this review was not the major question of those projects. These studies have been reviewed and analyzed.

Schedule for Completion

A draft report has been prepared.

Title: New Measures for College Admissions

Project/Job Number: 142/34   Project Director: Hunter M. Breland

Purpose

The purpose of this project is to review the possibilities for use and to evaluate new kinds of measures for college admissions that might supplement traditional indicators. A number of supplemental measures, including structured interviews and biographical data, are to be considered with respect to their reliabilities, validities, probable impact, cost, and other practical problems.

Progress (as of August, 1979)

This project has been subsumed under a larger project, Personal Qualities in Admissions (jointly sponsored by CB and ETS).

Schedule for Completion

Project will be completed as part of the Personal Qualities Project.

Title: An Investigation of Persistence in Higher Education, Phase I

Project/Job Number: 142/35 **Project Director:** Thomas L. Hilton

Purpose (Phase I)

1) To identify the major types of persisters according to the pattern and duration of attendance in various institutions of higher education.

2) To compare "high-potential" dropouts who persisted with those who did not.

3) To conceptualize a comprehensive model of the persistence process and test by means of NLS data file, including 2nd follow-up data.

Progress (as of August, 1979)

All the necessary analyses have been completed, and a draft report has been completed. This draft is now being revised.

Schedule for Completion

December, 1979.

Title: Career Skills Assessment Program Validity Study

Project/Job Number: 142/41 Project Directors: Jerilee Grandy
Charles Werts

Purpose

This study was designed to investigate the construct validity of three College Board Career Skills instruments: Career Awareness, Career Decision Making, and Employment Seeking Skills. A sample of 1500 high school students took two of the three Career Skills measures (the three were spiralled in pairs), the PSAT, a Career Development Questionnaire (CDQ), and a Test of General Information (TGI). A confirmatory factor analysis was used to estimate the extent to which each of the Career Skills instruments is measuring something uniquely different from verbal and mathematical aptitude (discriminate validity) and the extent to which it is measuring the same kinds of skills as the CDQ (convergent validity). Results suggest that about one-third of the variance in the CSAP scores can be predicted from measures of academic ability. Convergent validity with the CDQ was not established. As a by-product, the CDQ may become a marketable product in its own right.

Progress (as of August, 1979)

A final draft report revision has been completed.

Schedule for Completion

The final draft report is to be submitted for approval in September 1979.

Title: Discriminant Validity Study of CSAP

Project/Job Number: 142/48 Project Director: Donald A. Rock
Nancy Petersen

Purpose

The purpose of the study is to perform a confirmatory analysis in an effort to verify the construct validity of the CSAP materials. More specifically, project directors are investigating: (1) the construct validity of each of the skill areas, (2) the reliability of measurement of each of the skill areas, (3) the interrelationship among the skill area constructs and (4) relationships between the skill area constructs and outside constructs as measured by standardized achievement tests.

Progress (as of August, 1979)

Data collection is nearing completion; data analysis will follow in the fall.

Schedule for Completion

January, 1980.

Title: Development of CSAP User Validity Models

Project/Job Number: 142/49 **Project Director:** Jerilee Grandy

Purpose

This study is being executed in two phases: (1) Designs were developed according to which users can conduct their own content validation studies. A discussion of these designs appears in the CSAP Handbook. (2) Several institutions are being contacted to encourage their undertaking a validation study. Results of their studies will be included as examples in a manual designed to guide schools in conducting their own content validation studies.

Progress (as of July, 1979)

Phase I is complete; a manual with examples of user validation studies is being written.

Schedule for Completion

Phase I was completed in July, 1978; phase II will end in August, 1979.

Title: Validation of the Descriptive Tests of Mathematics Skills

Project/Job Number: 142/50 **Project Director:** Brent Bridgeman

Purpose

The purpose of this study is to determine the validity of the Descriptive Tests of Mathematics Skills (DTMS) for placement purposes. Various aspects of the project include a content analysis of the DTMS, assessment of student attitudes toward the quality of placement decisions, concurrent and predictive correlations of the DTMS with indicators of success in math courses (e.g., grades, class rank, final exam scores), a study of gains on the DTMS from the beginning of a math course to the end, and a trait-treatment interaction study.

Progress (as of July, 1979)

DTMS tests were administered to about 11,000 students in 34 institutions during the Fall of 1978. Posttests were administered at the end of the Fall term in 10 of these institutions, and two institutions participated as a posttest-only school. Rosters with the math course success indicators have been received. Content analysis forms from 128 faculty members were received and are now being processed. Some data from the field was late in arriving. Analyses of these data are just beginning.

A school that could meet the requirements for a high quality trait-treatment interaction study could not be located. A crude approximation may still be possible in one institution.

Schedule for Completion

October, 1979.

Title: Pre-equating the Test of Standard Written English

Project/Job Number: 142/51 Project Director: Isaac I. Bejar

Purpose

To study the feasibility of pre-equating Test of Standard Written English forms using Item Characteristic Curve Theory.

Progress (as of July, 1979)

About half of the required LOGIST (item calibration) runs needed have been completed. Programming required to do analysis of covariance structure analysis of item type effect has also been completed. Analysis is now under way.

Findings

Preliminary analysis suggests that the three parameter logistic model fits TSWE data adequately. No results are available yet as to how well pre-equating is likely to work.

Schedule for Completion

December, 1979.

Title: Exploratory Analysis of Sex Differences in Response Date for the ATP Physics Achievement Test

Project/Job Number: 142/52 Project Directors: Patricia Wheeler
Abigail M. Harris

Purpose

Sizeable performance differences have existed between men and women on the ATP Physics Achievement Test for many years. The purpose of this exploratory study is to investigate two major questions, one focusing on the test itself and the second on the candidates:

1) Are there any systematic differences in male/female performance on individual items or subgroups of items which can account for or help us interpret the differences between males and females in overall performance on the test?

2) Are there any factors such as number of semesters of physics or math, or overall level performance on the ATP Physics Test which, when related to performance data on items can help us to interpret the overall performance differences and perhaps gain insights into the unique characteristics of the female physics student?

Progress (as of August, 1979)

A review draft has been completed. This draft is presently being reviewed.

Schedule for Completion

Upon receipt of comments from reviewers, a final report will be prepared.

Title: Older Students and the SAT

Project/Job Number: 142-53 **Project Director:** Gita Wilder

Purpose

The purpose of this study is to assess the validity of the SAT with respect to older students. A sample of about 1,000 students who were 21 or older at the time of application for admission to college will be selected for study in an effort to discover how well the SAT predicts subsequent academic performance. A subsample of the students will be interviewed to learn something about the admission process for older students.

Progress (as of August, 1979)

The major efforts of the project to date have been directed toward identifying schools with significant numbers of older students who have taken the SAT. (Because many schools exempt older students from the testing aspects of the admissions process, finding an adequate sample is one of the most difficult tasks of the entire endeavor.) CEEB regional office personnel have provided contacts with admissions officers in their respective areas.

Schedule for Completion

It is intended that the sample be identified by early fall 1979 so that the collection of data may be carried out during the early part of the 1979-80 academic year.

Title: Supplementary Analyses of Data Obtained for the Study of "Taking the SAT"

Project/Job: 142-54 Project Director: Donald E. Powers

Purpose

The objectives of the study are to conduct further analyses of the survey data obtained for the study of Taking the SAT: A Test Familiarization Booklet to answer the following questions:

1. In what ways do students prepare for the SAT?
2. What are the relative effects on SAT scores associated with various types of preparation?
3. Were there any effects of using Taking the SAT that were not reflected in test scores (e.g., rate of work and guessing behavior)?

Progress (as of August 1979)

A draft report has been prepared describing the incidence and patterns of use of various modes of preparation for the SAT in a national random sample of candidates. PSAT scores are now being retrieved from score files for matching with SAT scores and other demographic variables. These data, along with students' reports of preparation activities, will be used to assess the relative effects of various methods of preparation.

Schedule for Completion

Two additional reports are scheduled for completion by December 31, 1979.

Title: Feasibility of Studying SAT Validity for Mexican-American Students

Project/Job Number: 142/55 **Project Director:** Gerilee Grandy

Purpose

The purpose of the feasibility study is to establish whether a sufficient number of institutions with large Mexican-American enrollments can be found and recruited to participate in a validity study. If so, a two-year study of the relationship between SAT and college grades will be conducted to determine whether the low correlations between SAT and college grades for Mexican-Americans (as found in earlier studies) can be attributed to (1) lesser intrinsic relationship between SAT and freshman GPA, (2) lower reliability of the SAT, or (3) lower reliability of the GPA for Mexican-Americans, or (4) some combination of these factors.

Progress (as of August, 1979)

The project has just begun.

Schedule for Completion

The feasibility study will be completed by September 30, 1979.

Title: Handicapped Students and the SAT

Project/Job Number: 142/56 Project Director: Marjorie Ragosta

Purpose

The purpose of this study is a preliminary investigation of (1) the perceptions of handicapped students about the SAT and the conditions under which it is administered especially as they relate to the individual's handicapping condition.

Progress (as of August, 1979)

The project is divided into two phases, the first of which has been completed. Phase I consisted of three tasks: (1) a literature search designed to look for optimal testing conditions for persons with various handicaps, (2) a telephone survey designed to locate colleges and universities using the SAT and having large populations of handicapped students, and (3) development and field testing of an interview schedule for use in Phase II of the study.

A progress report was prepared for approval before Phase II would begin. Phase II includes interviews with handicapped students in six selected universities and the production of a final report.

Schedule for Completion

Phase II is scheduled for completion December 31, 1979.

Title: Personal Qualities in Admissions - Phase I

Project/Job Number: 840/XX Project Director: Warren W. Willingham

Purpose

The primary purpose of the project is to evaluate the usefulness of personal qualities (beyond the usual grades and test scores) in selecting college students and in understanding student decisions to attend college. Nine private colleges are collaborating in the study.

Progress (as of July, 1979)

Considerable staff time has been devoted in the early stages of the work to planning meetings, organizational activities, visits to campuses of participating colleges, and other activities essential to establishing good working relationships and getting underway. Similarly, institutional participants faced major tasks early in the project--preparing a detailed description of their admission policy and procedures, adjusting their operational routines to the demands of the project, and training staff for the work involved.

Institutional work on behalf of the project has been substantial. Each of more than 25,000 admissions folders has been given a careful reading by admissions staff and faculty in order to produce some 20 special ratings of personal qualities. At each institution a number of records have been rated two or more times independently to provide information about the consistency of different types of judgments. Finally, those ratings and other essential parts of the application--some 300,000 sheets--are being copied, assembled, and forwarded to the project staff in Princeton. All of this work has proceeded very nearly on schedule with no serious mishap.

The initial stage of data coding and entry on tape has been completed for approximately half of the total group. A second major stage of data preparation involves systematic rating by ETS staff of a number of personal qualities of students (e.g., work experience, leadership, community activities, special talents, references, etc.). Considerable progress has been made in experimenting with subjective rating procedures and training project staff to make reliable research ratings of various aspects of the applications that will complement the judgments of institutional staff and faculty. That experimental work has proceeded quite satisfactorily and it is now clear that raters can be trained to distinguish reliably among a number of important student qualities.

Schedule for Completion

Phase I of the project is scheduled for completion in January 1980.

B. PROGRESS REPORTS FOR RESEARCH PROJECTS OF THE

GRADUATE MANAGEMENT ADMISSION COUNCIL

GMAC

Title: Assessment of Interpersonal Competence

Project/Job Number: 693/77 Project Director: Lawrence J. Stricker

Purpose

The aim of this study is to (a) develop a prototype measure of interpersonal competence based on the videotape presentation of a common social situation in a business setting and (b) obtain preliminary data on the device's item functioning and reliability, and, to the extent feasible, validity.

Progress (as of July, 1979)

As a first step in detailed planning of the instrument, authorities on interpersonal processes, executive assessment and training, and simulation measures were extensively consulted. The pertinent literature was also reviewed. Based on the consultations and review, a blueprint for the measure was prepared. A large number of scenes portraying a variety of situations were drafted and edited. The scenes and accompanying instructions, in paper and pencil format, were tried out and successively revised after a series of pilot tests with individuals and small groups at ETS as well as with two classes of undergraduates at business school. A videotape containing all the necessary instructions together with exercises based on the scenes was filmed, narrated, edited, and copied. The videotape and accompanying material were pilot tested individually with ETS employees. A small scale study, in conjunction with a separate investigation by Wanda Rapacznski, was carried out. The videotape and other instruments, including experimental measures of accuracy in social perception, a vocabulary test, the Embedded Figures Test, and a background questionnaire, were given to 59 college-age women. Procedures for rating the responses to the videotape measure were developed, the cooperation of two expert judges obtained, and the data analysis is underway.

Schedule for Completion

September, 1979.

Title: Technical Report on Test Development and Score Interpretation for GMAT Users

Project/Job Number: 693/79 Project Director: William B. Schrader

Purpose

To provide GMAT test users and others concerned with the technical quality of the test and related services with the basic information needed for understanding the processes involved in test development and score interpretation. Reliability, validity, speededness, and score equating will be discussed in terms of their contribution to program objectives.

Progress (as of July, 1979)

Basic data and reference sources for the manual have been collected, and a draft of the initial section on test development has been written.

Anticipated Completion and Dissemination

Referenced work is substantially completed, and it is anticipated that it will be published in the fall of 1979.

Title: Differences in Item Performance Across Groups

Project/Job Number: 693-81 Project Director: Loraine Sinnott

Purpose

The study will examine the extent of item bias within the GMAT relative to groups characterized by their sex, race, age, or language fluency.

Progress (as of July, 1979)

The January 1978 test administration will provide the study's data. The sample has been drawn and data analysis is underway.

Schedule for Completion

It is anticipated that data analysis will be completed by August 30. A draft of the final report will be available for the Selection Committee's fall meeting.

Title: Construct Validity of the GMAT: A Factor Analytic Study

Project/Job Number: 693/82 Project Directors: Spencer S. Swinton
Donald A. Rock
Donald E. Powers

Purpose

A factor analytic study of the GMAT will be undertaken to determine relevant dimensions of test variation in several forms of the GMAT and for several subgroups. The relationship of each of several experimental item types to the dimensions underlying performance on current operational forms will also be determined.

Progress (as of August, 1979)

Specifications for drawing several samples from two GMAT administrations (from three test forms) have been developed. Samples are being drawn of 1,000 cases of item-level responses of examinees from forms ZBS1 (November 1977 and March 1978), ZBS2 and K-YBS1 (November 1977).

Schedule for Completion

All analyses are scheduled for completion by Spring 1980 so that a final report can be prepared by June, 1980. A detailed progress report dealing with structure and parallelism of the forms is to be submitted on November 30, 1979.

C. PROGRESS REPORTS FOR RESEARCH PROJECTS OF THE

GRADUATE RECORD EXAMINATIONS BOARD

GRE

Title: Test Appropriateness Project

GRE Number: 75-3 Project Director: Michael V. Levine
Project/Job Number: 822/01

Purpose

Current aptitude test models and scoring procedures may be inappropriate for some exceptionally creative examinees, examinees with atypical cultural or educational backgrounds, and copiers. Some of these aberrant examinees are expected to produce unusual patterns of item responses. The main purpose of this project is to develop an objective way to use the examinee's pattern of right and wrong answers to decide whether a test is an appropriate measure of his or her ability.

Findings

Several appropriateness indices (statistics for identifying aberrant candidates) have been defined and evaluated. The basic idea of appropriateness measurement has been supported by Monte Carlo studies and studies with GRE data; moderately and extremely aberrant candidates can be identified by their pattern of item scores.

A Research Bulletin (RB-76-31) has been completed and submitted to the Journal of the American Statistical Society. Two additional articles are being prepared. Work on the project is continuing at the University of Illinois, where a dissertation on the topic has been completed.

Schedule for Completion

A final project report will be submitted to the GREB Research Committee in September, 1979.

---

*This project is sponsored jointly by CEEB, GREB, and ETS.

Title: Norming and Validation of a Machine-Scorable Test of Field Dependence-Independence

GRE Number: 75-7 Project Directors: Joel Campbell
Project/Job Number: 546/24 Thomas Donlon

Purpose

This project undertook an experimental administration of a machine-scorable test of field dependence-independence in the October 1976 GRE Aptitude Test administration to a sample of about 20,000 students. Using data from the background questionnaire in the registration form and follow-up questionnaires immediately after the test and approximately 20 months later, test results will be related to undergraduate major, graduate major, persistence in major chosen, and academic performance.

Progress (as of July, 1979)

Follow-up questionnaires were mailed in November 1976 to all students registered at those centers where the Figure Location Test had been administered in October. A second mailing was made in April 1977 to non-respondents. An initial analysis of approximately 9,500 cases was reported to the research committee in January, 1978.

A second questionnaire, designed to elicit graduate school experience as well as some attitudes and values was mailed in early July 1978. The follow-up, however, was delayed until October by processing delays, such as the necessity of hand-darkening the machine-printed registration numbers which were rejected in substantial number by the scanner. A total of 5,378 usable questionnaires were returned.

Data analysis for the validation part of the study has been started. Initial findings show limited evidence for any general contribution by FLT to graduate performance beyond what is already available through GRE V and Q. More detailed analysis by major field is continuing.

Schedule for Completion

Results of the first phase of the project were reported to the Research Committee in January, 1978. A final report will be submitted to the GRE Board in January, 1980.

Title: Career Achievement and GRE Scores

GRE Number: 76-1 Project Director: William B. Schrader
Project/Job Number: 546/25

Purpose

To examine the relationship between GRE scores and the scientific productivity of psychologists and historians, and to explore the usefulness of biographical directories and citation indices as sources of information on career attainments.

Findings

Among the various measures of productivity and professional contribution included in the studies, the total number of citations, as obtained from the Social Sciences Citation Index, was considered to be of special interest. Correlations of the three GRE scores with total citations for 150 psychologists and 83 historians were as follows:

	Psychologists	Historians
GRE-Verbal	.27	.18
GRE-Quantitative	.31	.00
GRE Advanced	.41	.14

Although all three coefficients for psychologists were significant at the 1% level, and none of the three coefficients for historians were statistically significant, the samples were too small to warrant detailed comparisons or strong conclusions. The results are sufficiently promising, however, to suggest that further studies based on larger samples would provide useful information about the extent to which career achievement in various fields can be predicted from GRE scores.

Dissemination

The report for psychologists was issued as GRE Board Research Report GREB No. 76-1R, September 1978. A final draft of the report on historians was submitted to the GRE Board in April, 1979, and will be published in the GRE Board research series.

Title: The Effect of Increased Test-Taking Time on Test Scores by Ethnic Group, Sex, and Age

GRE Number: 76-6 Project Director: Cheryl Wild
Project/Job Number: 546/30

Purpose

The project is designed to determine whether the allotment of more time per test item will result in a differential score improvement for selected subgroups of test takers. Experimental pretests with the same sets of verbal and quantitative items were administered with 20- and 30-minute time limits at selected test centers.

Findings

This study investigates the effects of increasing the test time to reduce the speededness of the verbal and quantitative experimental sections of the Graduate Record Examinations (GRE) Aptitude Test. In December 1976, at approximately 550 domestic test centers, 20- and 30-minute versions of a verbal experimental test and of a quantitative experimental test were administered with the operational test. The effects of testing time on scores for minority-group versus majority-group examinees, on male versus female examinees, and on examinees by years elapsed since they received the baccalaureate degree are investigated. Although the 20-minute experimental tests are generally more speeded than the 30-minute tests, the ten additional minutes resulted in a small score gain for all groups, and differential score gains were not found between the subgroups. On the basis of the results of this study, the extension of testing time for the GRE Aptitude Test for the purpose of reducing intergroup differences is not indicated.

Schedule for Completion

Submitted to the GRE Board in April, 1979 and published in July 1979, as GREB Research Report No. 76-6R.

Title: Older Students and the Graduate Record Examinations: A Preliminary Investigation

GRE Number: 76-13
Project/Job Number: 546/64

Project Directors: Joan Baratz
Terry Hartle

Purpose

The purpose of the study is a preliminary investigation of the characteristics and graduate student plans of older persons who took the GRE in a recent year, the problems encountered by a sample of graduate programs in the admission of older students, and recommendations of ways in which the Board and program staff might assist universities in dealing with the admissions problems presented by older students.

Findings

The records of more than 230,000 men and women who took the GRE Aptitude Test during 1975-76 were examined for differences between those who were completing college at the usual age of 22 or younger (about 48% of the test-takers), and those who had completed their undergraduate degrees more than 8 years earlier (about 8% of the test-takers.) Though on the average the older groups scored lower on the verbal and quantitative GRE tests, more of the differences in verbal scores were consistent with differences in anticipated fields of graduate study. The average verbal scores of older men and women usually were about the same as those of 22-year olds when comparisons were made within a given field. In fact, women more than 8 years beyond the bachelor's degree often averaged higher verbal test scores than did younger test-takers who were planning graduate study in the same field.

Older test-takers consistently averaged lower scores on the quantitative test. Differences in performance on the verbal and quantitative tests may reflect differences in opportunities for adults to maintain verbal and quantitative skills, with much more verbal stimulation in everyday living for most persons. Also, proportionately fewer older candidates planned graduate study in scientific or technical fields.

Dissemination

A final report on the backgrounds of older GRE test-takers was completed in April, 1979, and is being published as GRE Board Research Report No.76-13R.

Interviews also were conducted on about a dozen campuses concerning departmental experiences with older applicants for graduate admission. A report of these interviews, with recommendations to the GRE Board about the uses of GRE scores, was submitted to the GREB Research Committee in September, 1978.

Title: Further Research on Data Collected for the Dimensions of Quality Project

GRE Number: 77-2 Project Director: Mary Jo Clark
Project/Job Number: 546/57

Purpose

The purpose of this work is to increase our understanding of the processes and outcomes of doctoral education through the intensive study of selected graduate programs. Data were collected from advanced graduate students, faculty members, and recent alumni in departments of chemistry, history, and psychology at 25 representative universities for the recently completed CGS/ETS project on the Dimensions of Quality in Doctoral Education. Questions such as the effect of subdisciplines on program evaluation, the employment patterns of graduates, and personal vs. situational influences on faculty productivity will be investigated by using individual questionnaire respondents as the unit of analysis.

Progress (as of July, 1979)

The analysis of the extent to which questionnaire responses of faculty members and students can be attributed to consistent individual, university, and department effects is being completed in a way that will be most useful in the revision of questionnaires for operational use in program reviews (currently sponsored by GRE and CGS), as well as for a report to the GRE Board. Analyses have also been carried out to investigate sex differences in faculty opinions and behaviors, faculty publication records in relation to the prestige of their employing departments, and variations in questionnaire responses by faculty members in the subdisciplines of psychology. Draft articles for publication are underway.

Schedule for Completion

Findings will be disseminated as articles in appropriate professional journals, with a final report to the Board as soon as the articles are completed.

Title: Information Needs of Prospective Graduate Students

GRE Number: 77-8 Project Director: Rodney T. Hartnett
Project/Job Number: 546/32

Purpose

The purpose of this research is to learn more about three separate but related aspects of the school-to-applicant flow of information at the graduate school level: (1) What factors are important influencers to applicants' choices of specific graduate programs?; (2) What specific information about graduate departments would probably be most helpful to prospective graduate students and how likely is it that such information can be obtained?; and (3) What would seem to be the most useful and efficient procedure for collecting, summarizing, and reporting such information to prospective students?

Findings

The great majority of graduate students interviewed reported general satisfaction with the information available to them about graduate programs. In this sense, then, there does not appear to be an information need among prospective graduate students. On the other hand, there is some indication that prospective students could have benefitted from more and better information if they had known more about various aspects of graduate education prior to the time they enrolled. It is concluded that more information could improve the graduate school choices of some students, but that it is not likely to do so without some formal effort to assist prospective students clarify their own interests and goals and identify the types of information that they need to have to enable an accurate estimate of the appropriateness of a given department for them.

Schedule for Completion and Dissemination

A draft final report for the fall, 1979, Board meeting.

Title: Construct Validity of the GRE Aptitude Tests Across Populations: An Empirical Confirmatory Study

GRE Number: 78-1 Project Director: Donald A. Rock
Project/Job Number: 546/50

Purpose

The study will examine the GRE Aptitude Tests for psychometric bias across populations by statistically (1) verifying the a priori factor structure of the GRE-V, GRE-Q, and GRE-A; (2) testing the goodness of fit of this structure across black, white, male and female populations; (3) testing if the metric or scale units of the test scores are the same over these populations; and (4) testing whether the a priori test reliabilities are equal across populations. Subscales of item types will be examined in the same way.

Progress (as of July, 1979)

The analysis of the GRE-V, GRE-Q, and GRE-A indicated support for the hypothesis that each of the a priori factors was measuring the same thing for white and black men and women. The measures also were in the same units with equivalent accuracy as indicated by both reliabilities and the standard errors of measurement. Further study, particularly to investigate the characteristics of groups through the use of extension variables is underway.

Schedule for Completion

A preliminary report in September, 1979, with a final draft report in January, 1980.

Title: A Study of the Effects of Special Preparation on GRE Analytical Scores

GRE Number: 78-2 Project Directors: Spencer S. Swinton
Project/Job Number: 546/61 Donald E. Powers

Purpose

The primary purpose of the project is to determine the susceptibility of analytic item types to short-term instruction by developing coaching materials and pilot testing them at one or two local universities. If warranted by the results of this feasibility study, a full scale study using the coaching materials will be proposed as a separate study.

Progress (as of July, 1979)

Existing test-taking manuals were secured and examined. The project directors have also gained some familiarity with the type of preparation for the GRE analytical section that is currently being offered by proprietary coaching schools. A number of practice items have been written. Data to test the similarity of these items to those included in the GRE sample test resulted in reliabilities of .92 and .87, suggesting that the new sample items measure skills quite comparable to those measured by the items from the sample test. The test-taking strategies of several subjects have been reviewed in order to analyze typical errors in each item type. Familiarization materials have been developed and arrangements for pilot use of these materials has been made at a local college. A proposal for packaging these materials in a manner suitable for mail delivery and evaluation (after they have been pilot tested) has been submitted to the GRE Research Committee.

Schedule for Completion

An extended progress report is being submitted to the GREB in September, 1979, with a draft final report expected in January, 1980.

Title: Documented Accomplishments: Phase III

GRE Number: 78-3 Project Directors: Leonard Baird
Project/Job Number: 546/62 Joan Knapp

Purpose

The project is the third stage of a four-stage project designed to develop an inventory of documented accomplishments for graduate selection. (See GRE project No. 77-3 for a report of the earlier work.) This phase of the study is designed to tryout the materials developed in earlier stages in graduate departments. The major purposes of this study are to examine the instrument developed in Stage Two in terms of: (1) technical soundness, and (2) the feasibility of its use in the admissions process. This study represents the transition of the use of documented accomplishments to predict graduate school achievement from a research phase to a more operational phase during which the instrument will be tested in a real life situation.

Progress (as of July, 1979)

Entering students at 27 departments in 10 diverse universities were administered the inventory. The fields of biology, English and psychology were included. The same students were sent a followup questionnaire about their activities during their first year in graduate school at the end of their academic year. Both instruments are on tape and analyses are underway. Substantive analyses of students' free responses to the inventory are also being conducted.

Schedule for Completion

A draft final report to the GREB is expected in January, 1980.

Title: Content Validity of Selected GRE Advanced Tests

GRE Number: 78-4 Project Director: Miles McPeek
Project/Job Number: 546/63

Purpose

The purpose of the study is to obtain information about the correspondence between GRE Advanced Tests: Biology, Literature in English, and Political Science. The study will serve as a pilot test for succeeding content validity studies of the remaining Advanced Tests.

Progress (as of July, 1979)

A revised study design, reflecting the major changes in the direction of the study requested by the Research Committee, was approved by the Research Committee at their April meeting. Development of the final form of the questionnaires to be used for each discipline is underway.

Schedule for Completion

A draft final report will be completed by June, 1980.

Title: Analysis of the Data Collected from GRE Test-Takers, 1975-1977

GRE Number: 78-5 Project Director: Paul Holland
Project/Job Number: 546/04

Purpose

This project will examine summaries of background information supplied by GRE test-takers and published by the GRE Board (Altman and Holland, 1976; Altman, 1977). Modern data analytic techniques will be applied to improve the quality of the information, simplify its presentation, and make it more useful to policy makers and researchers. The results will be used to improve future summary reports of similar data.

Findings (as of July, 1979)

This study includes five sub-studies and separate summary reports are being prepared for each. The general conclusions of these studies are as follows: (1) Standardizing ethnic group mean GRE scores on background variables is feasible and does have a small but clear effect. It was recommended that similar standardizations be done using the new SES data and the results of this new analysis are now available in a preliminary form. (2) The patterns of missing data in the background data are very numerous but are almost completely explainable by the operation of two nearly independent phenomena, (a) answering all but one question, or (b) quitting after answering some of the questions in sequence. (3) The effect on missing data of the currently used criterion for inclusion in the published tables is of medium importance and it should be changed to reflect the instructions given to the respondees. (4) The analysis of the published tables of counts and percentages using log-linear models revealed a number of ways in which complex tables could be simplified without losing essential information. In addition, the stability over time of several aspects of the test-taking population was detected. (5) The analyses of the published tables of GRE means revealed that many can be described more simply by row and column effects. Often, the interactions are not large and do not justify elaborate disaggregated tables.

Schedule for Completion

A draft final report to the GRE Board is expected in January, 1980.

Title: A Study of the Correlational Validity of the Restructured GRE Aptitude Test

GRE Number: 78-6 Project Director: Kenneth Wilson
Project/Job Number: 546/65

Purpose

The project will study the correlational validity of the restructured GRE Aptitude Test (shortened tests of verbal and quantitative ability and new test of analytical ability) by evaluating scores on the GRE Aptitude Tests administered during 1977-78 in relation to first year graduate school grades or other performance criteria in a sample of graduate departments in each of about 10 disciplines. Particularly, the study asks: does the GRE-A score increase the ability to predict typical performance criteria in a variety of fields? If so, in which fields does the information provided by GRE-A appear to supplement GRE-V and/or GRE-Q, and in what ways?

The research will: (a) evaluate trends in patterns of zero-order coefficients of predictors with criteria; (b) pool within-group data across departments within fields to explore the relative weighting of the respective Aptitude Test scores; (c) assess the correlational validity of a residual score reflecting variation in GRE-A independent of V and Q; and (d) for selected institutions, regress the accept/reject decision on selected admissions variables.

Progress (as of July, 1979)

As of July 1979, 39 graduate schools represented by a total of over 150 departments (a) have provided names of students eligible for inclusion in a study sample and (b) are in the process of providing a first-year graduate grade point average for these individuals. Many of the departmental samples are extremely small (i.e., less than 10 cases) but it is expected that pooling procedures will yield interpretable and useful results. In addition to providing information about enrolled first-time graduate students, several graduate schools also have identified GRE score senders (a) who applied but were not accepted, and (b) who were accepted but did not enroll. Using these data, analyses will be made of differences between various admissions stages (e.g., accepted vs. rejected, accepted enrolled vs. accepted nonenrolled) with respect to GRE scores and other data. All work on this project is being coordinated with the GRE program's development and implementation of a Validity Study Service for graduate departments.

Schedule for Completion

A final report will be submitted to the GRE Board in September, 1980.

Title: Item Indexes of Differential Subgroup Performance

GRE Number: 78-7 Project Director: Lawrence J. Stricker
Project/Job Number: 546/66

Purpose

One promising approach to minimizing test bias is to identify, during the course of the test construction, items that perform differently for the relevant subgroups. This study will explore the usefulness of a new index in identifying differential performance of GRE Aptitude Test items with respect to race (white and black) and sex: the partial correlation between the item and subgroup total score being held constant. This procedure is free of the problems associated with most existing methods and is applicable to comparatively small samples. The investigation will (a) determine whether differentially functioning items can be identified by this index; (b) assess whether these items are the same as those earmarked by two of the most important other methods of studying item bias--item difficulty plots and item characteristic curves; and (c) attempt to ascertain content characteristics that distinguish the items uncovered by the various procedures. Subanalyses will also evaluate the extent to which the results (a) are altered when the race and sex samples are matched on the various GRE Aptitude Test scores, and (b) can be accounted for by chance, as judged by Monte Carlo analyses of the smallest race and sex samples.

Progress (as of July, 1979)

Data for the GRE Aptitude Test (ZG42) used in the October, 1977 and January, 1978 administrations were analyzed. Samples were drawn of one-fourth of the white and all the black examinees with these characteristics: never attended graduate school full-time, currently a senior or unenrolled college graduate, undergraduate major in social sciences, no test irregularity, never took the GRE previously and native of the United States. The four basic samples were white males (N = 1122), white females (N = 1471), black males (N = 284), and black females (N = 626). Samples with matching true score distributions (e.g., matched on scores for odd items on the Verbal test for analyses of even items on that test) were also obtained; the matched samples were white females/black females, and white males/white females. Monte Carlo samples were also secured: a random sample of white males, the same size as the basic sample of black males; and another random sample of white males, identical in size to the basic sample of black females. The analyses of the partial correlations have just been completed and are being examined in detail; the analyses of the other indexes are underway.

A substantial proportion of the item partial-correlations were significant ($p < .05$) in the main analyses for race and sex. However, few of the correlations exceeded .10 in absolute size and the significant correlations varied in their direction, not predominantly favoring one subgroup. A detailed appraisal of these results is being made, focusing particularly on their consistency across samples and analyses.

Schedule for Completion

A final report is scheduled for completion by January, 1980.

Title: Research on a Revised Format of Analysis of Explanations Items

GRE Number: 79-1
Project/Job Number: 546/68

Project Directors: Sydell Carlton
Miles McPeek

Purpose

A factor analysis of the restructured GRE Aptitude Test suggested that the correlation between the Verbal and Analytical scores might be reduced by eliminating one of the response options associated with Analysis of Explanations items and replacing items keyed to this response with items keyed to one of the remaining four responses. The purpose of this project is to determine whether use of 4-option rather than 5-option Analysis of Explanations items in the Analytical section of the GRE Aptitude Test would result in Analytical scores that are more independent of Verbal scores while being equally reliable and valid.

Progress (as of July, 1979)

Two pretests, each containing the same four sets of previously tested Analysis of Explanations items, were prepared. In one pretest, the five-choice control format, the items remain as they were in final form administration. In the other, the four-choice experimental pretest, items with an A key have been removed and replaced by pretested items with B, C, or D keys but with similar pretest statistics. These two pretests will be spiralled in the October 1979 administration of the GRE Aptitude Test.

Schedule for Completion

Administration of four-choice and five-choice formats in October 1979; draft final report in April 1980.

D. PROGRESS REPORTS FOR RESEARCH PROJECTS OF THE

LAW SCHOOL ADMISSION COUNCIL

Title: Research on Proficiency in Legal Practice - Phase II, Study 2: "The Activities and Tasks of Lawyers in Litigation Practice"

Project/Job Number: 688/12   Project Director: Robert F. Boldt

Purpose

This study is concerned with obtaining a detailed and systematic description of the litigation-related work activities of the profession. The objective of the study is to gather quantitative ratings on the activities of lawyers in advocacy/trial practice, on the importance of these activities, as judged by litigation lawyers, to the discharge of their professional responsibilities, and to gather questionnaire information on demographic factors and the lawyer's characterizations of their practices. These data will then be analyzed to (1) identify groups of lawyers with similar perceptions of the importance of activities, (2) identify groups of lawyers that distribute time to these activities similarly, (3) examine the similarity of groups determined by time and importance, (4) identify those activities that, for the groups, are most frequent and most important, (5) relate group-membership to demographic data and descriptions of the types of practice, and (6) make recommendations for the content of rating forms that could be used in the evaluation of lawyer performance. The results will provide a description of the important tasks of lawyers in a form that will be useful in guiding the development of measures of the proficiency of lawyers in this segment of practice.

Progress (as of July, 1979)

A job analysis involving observation, interviews, and collection of time-log data for a small sample of attorneys has been conducted. A list of the activities of litigators has been compiled and reviewed. A background questionnaire and instructions for responding to the activities list has been developed, pretested, and revised.

Data collection materials were sent to a sample of 6,272 attorneys drawn from the 1976 Martindale-Hubble Law Directory and 1,059 women attorneys drawn from the 1974 Membership Directory of the National Association of Women Lawyers. Preliminary analysis indicates that after three followups 1,704 sets of ratings are available for the identification of groups and group members; and the analysis has been planned and initiated for the about 1,200 complete data cases.

Schedule for Completion

This study is scheduled for completion during 1979.

Title: A Study of the Characteristics of Law School Examinations Written by Minority and Non-Minority Law Students

Project/Job Number: 890/21   Project Director: Franklin R. Evans

Purpose

The study will attempt to identify ways in which groups of minority and non-minority law students systematically differ in their performance on a law school essay examination. A secondary purpose is to use the data to examine possible "bias" in law school grading and the differential prediction of an "unbiased" criterion.

Progress (as of July, 1979)

Nine law professors at eight law schools have participated by providing data from their TORTS classes. A common essay question was included in the TORTS final examinations. A twenty item multiple choice test in TORTS was also administered by 8 of the 9 professors. Supplementary information, such as undergraduate grades and first-year law school grades has been collected. The essay answers have been scored for several indices of readability, legal analysis, and overall writing quality.

Schedule for Completion

A final report will be available in 1979.

Title: Models of Law School Admissions Procedures

Project/Job Number: 890/22 Project Director: Franklin R. Evans

Purpose

The purpose of this project is to develop several models of admissions procedures. Each model will be evaluated by comparing predicted decisions obtained using the models with actual admissions decisions. Models will also be used to investigate factors such as cost, the values attached to certain decisions, and efficiency of the admissions process.

Progress (as of July, 1979)

During 1976 site visits were conducted to observe the admissions process at nine law schools. Quantitative data obtained from a subset of the schools are being analyzed. At the same time, brief narratives describing each law school's admissions procedures are being written based on information obtained during the site visits. A preliminary analysis of the data from one school is completed.

Schedule for Completion

A descriptive report of admissions procedures at the law schools is now in preparation.

Title: A Comparison of the Effectiveness of Several Methods of Using Adjusted Undergraduate Grades for Predicting First Year Law School Performance

Project/Job Number: 890/41   Project Director: Donald A. Rock
Franklin R. Evans

Purpose

This study is designed to determine if any of several methods currently used to adjust undergraduate grades at one or more law schools are effective across a representative sample of law schools.

Progress (as of July, 1979)

A draft report was submitted to the T D & R committee at its October, 1978 meeting. At that time the committee requested additional analyses. A revised draft report was prepared for the spring 1979 meeting. At that time the Committee requested further analyses. These are in progress.

Schedule for completion

Final report is scheduled for completion during winter, 1979.

Title: Validity Study of Three Experimental Item Types - Issues and Facts, Analytical Reasoning, and Evaluation of Evidence - Administered in 1975-76

Project/Job Number: 890/61   Project Directors: Miles McPeek

Purpose

This project is designed to study the validity of Issues and Facts, Analytical Reasoning, and Evaluation of Evidence for predicting first-year average grades in law school. Of major interest is the validity of combinations of these items with those currently in use in the LSAT (Logical Reasoning, Practical Judgment, Data Interpretation and Quantitative Comparison, Principals and Cases, Sentence Correction and Error Recognition). The effect on validity of substituting the new item types for certain of the current ones will also be evaluated.

Progress (as of July, 1979)

The three experimental item types were included in test forms administered in 1975-76. Criterion data were collected in the 1977-78 Validity Study Service.

At the request of the LSAC these analyses were postponed in order to collect criterion data on four additional new item types (analyses of explanations, supporting conclusions, deductive reasoning, and evaluating summaries) administered during 1976-77. These additional data will be collected in the 1978-79 Validity Study Service.

Schedule for Completion

Data processing and statistical analysis are scheduled for the summer of 1979. Results should be available during spring, 1980.

Title: LSAT in England

Project/Job Number: 890/67 Project Director: Barbara Pitcher

Purpose

The purpose of this study is to explore the appropriateness of using the Law School Admission Test for selection of law students in Great Britain and to study the effectiveness of the test for predicting grades in the first and subsequent years in British law schools.

Progress (as of July, 1979)

Five law schools in Great Britain agreed to administer the LSAT to their entering classes in early October, 1975. An old form of the LSAT, of which there was a sufficient supply available and for which item and test analysis results were on file, was selected for administration.

Answer sheets were received from four of the law schools; the fifth found it impossible, because of changes in personnel and illness, to cooperate as originally planned. Thus, a total of 473 answer sheets have been scored, from four British law schools.

Criterion data, including grades for the entire three-year period are being collected.

Schedule for Completion

The final project report is scheduled to be completed in 1979.

Title: Predicting Law School Grades for Minority and Nonminority Students: Beyond First-Year Averages

Project/Job Number: 900/05 Project Director: Donald Powers

Purpose

This study is being conducted in order to investigate (1) the predictive and possible differential validity of standard preadmission measures (UGPA and LSAT scores) for forecasting the academic performance of minority and nonminority students beyond the first year of law school and (2) the predictability of criteria other than global averages (e.g., composites based on sets of courses).

Progress (as of July, 1979)

Procedures are now being devised for collecting data from law schools that provided information for a previously completed study and requests for data have been made to 30 of them. Sample transcripts have been obtained from about 30 law schools to assess the ways in which such transcripts may differ and to suggest procedures for their analysis.

Schedule for Completion

A draft final report will be prepared for the Spring 1980 meeting of the Test Development and Research Committee.

Title: Differential Predictive Validity of Writing Ability Scores for Minority Students

Project/Job Number: 900/06 Project Director: Donald Powers

Purpose

The purpose is to investigate the possibility that the Writing Ability section of the LSAT may be a better predictor of first-year law school performance for minority students than for nonminority students. The Test Development and Research Committee's interest in this issue was stimulated by data from a previous study suggesting this possibility. The present study will re-examine the previously collected data and attempt to replicate findings based on those data by using additional data.

Progress (as of February, 1979)

Reanalysis of previously collected data is partially completed. Results of selected schools from the current year's Validity Study Service have been obtained for comparison purposes.

Schedule for Completion

A final report will be prepared for the Spring 1980 meeting of the Test Development and Research Committee.

E. PROGRESS REPORTS FOR RESEARCH PROJECTS OF THE

TEST OF ENGLISH AS A FOREIGN LANGUAGE

TOEFL

Title: A Comparison of Non-native Speaker Performance on TOEFL and Verbal Aptitude Tests

Project/Job Number: 575/31 Project Director: Paul J. Angelis

Purpose

The purpose of this study was to relate the performance of non-native English speaking graduate and undergraduate applicants on TOEFL to their performance on the GRE verbal or SAT verbal, including TSWE. A secondary purpose was to gather assessments of the appropriateness of the item formats and test content from a group of English as a second language specialists.

Progress (as of July, 1979)

The final report is now in the review process.

Schedule for Completion

The final report will be available by late summer 1979.

Title: An Exploration of Speaking Proficiency Measures in the TOEFL Context

Project/Job Number: 575/32 Project Directors: John L.D. Clark
Spencer S. Swinton

Purpose

This project involves the development and experimental administration of a variety of item types and formats for the direct testing of English speaking proficiency within the TOEFL candidate population. In addition to the experimental measures, examinees will also be administered the Foreign Service Institute ("FSI") proficiency interview as a criterion measure of speaking ability, as well as the non-speaking (listening comprehension, structure, reading) sections of the current TOEFL. Factor analysis and other techniques will be used to determine the interrelationships between experimental, criterion, and TOEFL measures. These analyses, together with a consideration of the relative ease of administration, face validity, and other characteristics of the experimental item types, will be used to identify a subset of item types with potential for operational use within the TOEFL program as a direct test of English speaking proficiency.

Progress (as of August, 1979)

The final report has been completed in internal review form and is being circulated for review.

Schedule for Completion

Final printed report should be available by August 30, 1979.

Title: Language Skills in Academic Study and Performance on TOEFL

Project/Job Number: 575/60 **Project Director:** Paul J. Angelis

Purpose

The purpose of this study is to collect and analyze data indicating the English skills required for successful academic work in the principal graduate and undergraduate fields studied by foreign students in U.S. institutions. The principal components of the study will be collection of curricula information from institutions and interviews with students at the completion of a full year of study.

Progress (as of August, 1979)

A faculty questionnaire has been prepared and reviewed by a committee of representatives from the fields chosen (Business, Chemistry and Civil Engineering). Pilot and Full administrations of the questionnaire will be conducted in fall 1979. Graduate student interviews are being conducted in summer 1979.

Schedule for Completion

The study is planned for completion by early 1980.

Title: Development of a Manual for Institutional Language Growth Analysis

Project/Job Number: 575/61 Project Directors: Spencer S. Swinton
Paul J. Angelis

Purpose

This field-based development project will yield a manual to assist English-as-a-Second-Language departments in analyzing predicting, and interpreting changes in language test scores over time.

Progress (as of August, 1979)

A worked example, involving a pretest, a reliability test, and a posttest of data from the fall semester at San Francisco State University has been analyzed using the SPSS package and presented to members of the department. Initial reaction suggested that the original presentation was too technical, and a revision, in a more step-by-step format, is being prepared to guide department members in performing and interpreting the analyses on their own.

Schedule for Completion

The revised manual will be tried out with an additional department (using summer data in October). Final revisions will be made in November and December, 1979.

Title: Factor Analysis of the Test of English as a Foreign Language for Several Language Groups

Project/Job Number: 575/62 Project Directors: Spencer S. Swinton
Donald E. Powers

Purpose

This study has three purposes. (1) To obtain a clearer understanding of the psychological constructs underlying TOEFL, (2) To determine dimensions important to control in maintaining parallelism of test forms, (3) To ascertain the comparability of test dimensions for examinees of different native language groups.

Progress (as of August, 1979)

Item responses by samples of approximately 1,000 subjects in each language group have been drawn, and the matrices of tetrachoric correlations computed. An assessment of the impact of applying Carroll's correction for guessing yielded the conclusion that the procedure would increase the magnitude of the common factor but would not eliminate apparent "difficulty factors." Even and odd items were factored separately in each language group, with two interesting results: 1) Because of differing distributions of verb-form and modifier (adjective and adverb) items among the evens and odds, minor factors emerged from these analyses, factors which were not clearly isomorphic across the two halves of the test, and 2) Within even or odd item analyses, the factor structures are markedly different across language groups. This latter finding was not unexpected. Although listening items cohere across all language groups, structure, written expression, and reading items show different patterns of covariance in the several groups. Some of these patterns may be explained by differing distributions of cognates (e.g., "global" in Spanish) and false cognates (e.g., "relation" in Spanish), but two factors tentatively identified as "formal" and "colloquial" vocabulary seem to show only moderate overlap in specific items involved across groups.

Because of the lack of agreement of the even and odd half analyses, the full 150 by 150 tetrachoric matrices have been analyzed, and final interpretation awaits the results of these analyses.

Schedule for Completion

A draft final report is scheduled for completion by September 1979.

Title: TOEFL Item Performance Across Language Groups

Project/Job Number: 575/63 Project Director: Donald L. Alderman

Purpose

The purpose of this project is to identify TOEFL items that differ significantly in difficulty or discrimination across several language groups. Estimates of item difficulty and actual distractor choices will lead to the identification of items which exhibit different response patterns across language groups. These items will then be reviewed by linguistic experts in an attempt to establish guidelines for subsequent test development. Tentative recommendations would be tested in an analysis of another form of the test as a check on generalizations.

Progress (as of August, 1979)

Item analyses have been conducted for seven language groups with a duplicate sample drawn for one group. Item characteristic curves (three parameter logistic curves) prompted study of alternative approaches to identifying particular items for linguistic analysis, and of methods for presenting results across several groups in the absence of a natural reference group. Other indices of item difficulty confirmed variation in performance dependent on native language. Preliminary results have been shared with test development committees and will be subject to further study this fall in order to detect possible linguistic patterns in item difficulty.

Schedule for Completion

Another form of the test will be examined this summer and a report submitted in the winter.

Title: An Exploratory Study of a Collaborative Research Effort with the Institute of International Education (IIE)

Project/Job Number: 575/64 **Project Director:** Gita Wilder

Purpose

The purpose of this effort is to examine in detail the data housed at IIE representing about 2,500 Venezuelan students who have received grants to study in the United States. The files contain socioeconomic data, information about foreign language study and examinations, academic study in the United States, and secondary school background. The exploratory.effort was initiated in order to assess the completeness of the records and to assess the kinds of questions that might be asked of the data.

Progress (as of December, 1978)

A systematic review of the files has been carried out and a draft report has been prepared describing their contents. Some information from the files has been summarized and is available in the form of data sheets and a computer printout. Facsimiles of student transcripts representing a subset of the entire population have been made available to ETS.

Schedule for Completion

A draft report has been prepared indicating what information is available and for what subgroups of students. Some questions that might be posed of the data have been presented along with some proprosed approaches to answering the questions. The draft is presently under review by ETS and IIE staff.

Title: TOEFL Speaking Test Validation Study

Project/Job Number: 579/11 **Project Directors:** John L.D. Clark
Spencer S. Swinton

Purpose

The purpose of this study is to provide use-validity information for one major application of the new TOEFL Test of Spoken English (TSE): the prediction of the linguistic effectiveness, in classroom or other instructional situations, of teaching assistants or graduate instructors whose native language is not English but whose ability to communicate in English is crucial to their job performance. The validation procedure will involve comparing foreign teaching assistants' scores on TSE to their judged "communicative effectiveness" in instructional settings, as rated by native U.S. students taking their classes. The study will also provide for concurrent validation of the TSE as a measure of general speaking proficiency, using as a criterion the Foreign Service Institute (FSI) interview procedure.

Progress (as of August, 1979)

Since initiation of the project in June, initial contacts have been made with participating schools and arrangements made to train TSE scorers and FSI interview raters for the project. On-site testing at the institutions will be conducted in September and early October.

Schedule for Completion

Authorized project period is June 1, 1979 to August 31, 1980. However, project planning memorandum and schedule have been based on a target completion date of June 30, 1980.

Title: A Study of the Relationships Between Scores on the Graduate Management Admission Test (GMAT) and the Test of English as a Foreign Language (TOEFL)

Project/Job Number: 579/12 Project Director: Donald E. Powers

Purpose

The objective of the study is to prepare data illustrating the relationships among the TOEFL and GMAT scores and first-year averages of nonnative applicants to American graduate schools of management.

Progress (as of August 1979)

Requests for data pertaining to nonnative graduate management students and applicants have been made to admissions staff at nearly 30 graduate schools of management. Plans are also being developed to match the GMAT and TOEFL test scores of nonnative applicants.

Schedule for Completion

First stage analyses involving only GMAT and TOEFL scores are scheduled for completion by December 31, 1979. The second stage analyses using first-year averages will be completed by June 30, 1980 at which time a final report will be prepared.

Title: A Descriptive Analysis of Foreign Test-Candidates: Characteristics and Score Profiles on TOEFL and Major Admissions Tests

Project/Job Number: 579/13 Project Director: Kenneth M. Wilson

Purpose

To define, describe, and delineate the foreign candidate population served by TOEFL, and assess the extent and nature of overlap between TOEFL population and those served by GRE, GMAT, and ATP to analyze score profiles for the respective populations.

Progress (as of August, 1979)

Discussions with representatives of the admissions testing programs involved in the study have resulted in preliminary plans for extracting records of candidates who are not U. S. citizens (noncitizens) and U. S. citizens who indicate that they communicate best in a language other than English (nonfluent citizens) during the testing years 1977-78 and 1978-79, respectively. It appears that the most effective method for identifying "foreign" admissions-testing-program candidates is to use responses to a citizenship (background) question. Work required to establish code-equivalencies for variables such as country, major field, etc. across testing programs has been initiated. Inquiries have been undertaken internally regarding the problem of assessing linguistic distance--the possibility of classifying languages according to "distance" from English would facilitate analysis of score profiles.

Schedule for Completion

The project is scheduled for completion as of February 1981.

II. LISTINGS OF PROGRAM RESEARCH PROJECTS COMPLETED DURING THE LAST FIVE YEARS

College Board Research Projects Completed During the Last Five Years

Equating the scales of the Spanish-Language Prueba de Aptitud Academica and the English-Language Scholastic Aptitude Test of the College Entrance Examination Board	W. Angoff C. Modu RB-73-4
Learning style preferences among college students	R. Hartnett RB-73-14
College Level Examination Program: Its meaning to participants	P. Casserly PR-73-5
Validity of the quantitative comparison test	W. Schrader SR-73-60
Effect of moderate departures from parallelism on the precision of SAT equating	W. Schrader PR-73-41
Using self-reports to predict student performance	L. Baird CB Research Monograph #7
Prediction of college achievement among Mexican-American students in California	J. Warren RB-76-22 CB RDR 76-77-1
Verbal and mathematical ability of high school juniors in 1974: A norms study of PSAT/NMSQT	R. Jackson W. Schrader RB-76-27 CB RDR 76-77-2
Measuring the appropriateness of multiple-choice test scores	M. Levine D. Rubin RB-76-31 CB RDR 76-77-3
A study of college English placement and the Tests of Standard Written English	H. Breland PR-77-1 CB RDR 76-77-4
College placement and exemption workbook	L. Aleamoni (in press)
Item characteristic curve solutions to three intractable testing problems	G. Marco RB-77-11 CB RDR 76-77-5
Group comparisons for the Test of Standard Written English	H. Breland RB-77-15 CB RDR 77-78-1

No-need/merit awards: A survey of their use at four-year public and private colleges and universities	A. Sidar D. Potter 1978
Short-term instruction, testwiseness, and the Scholastic Aptitude Test: A literature review with research recommendations	L. Pike RB-78-2 CB RDR 77-78-2
Predicting long-term performance in colleges of minority and nonminority students: A comparative analysis in two collegiate settings	K. Wilson RB-78-6 CB RDR 77-78-3
College enrollment in the 1980s: Projections and possibilities	J. Centra RB-78-18 CB RDR 78-79-1
Population validity and college entrance measures	H. Breland RB-78-19 CB RDR 78-79-2
Multivariate talent flow analysis: A pilot study	K. Wilson RB-78-20 CB RDR 78-79-3
That's a very good question: The economic impact of credit by examination policies and practices	J. Valley 1978
The effects of special preparation on SAT-Verbal scores	D. Alderman D. Powers CB RDR 78-79-4
Construct validity of the SAT across populations: An empirical confirmatory study	D. Rock C. Werts CB RDR 78-79-5
The use, acceptance, and impact of Taking the SAT--a test familiarization booklet	D. Powers D. Alderman CB RDR 78-79-6

RB (Research Bulletin)
PR (Project Report)
SR (Statistical Report)
RDR (Research & Development Report)

GMAC Research Projects Completed During the Last Five Years

A study of the influence of undergraduate course content on the Admission Test for Graduate Study in Business as a predictor of success in graduate schools of business	F. Connelly W. Nord GMAC Brief #5
Indicators of college quality as predictors of success in graduate schools of business	B. Pitcher W. Schrader GMAC Brief #6
Predicting first-year average grades for interrupted and uninterrupted students in graduate business schools	B. Pitcher GMAC Brief #7
A survey of skills and abilities needed for graduate study in business	J. Campbell P. Casserly GMAC Brief #9
Criterion study, Phase IV: Career progress of MBAs six years after graduation	L. Crooks J. Campbell PR-74-8
Survey of women interested in management	G. Echternacht A. Hussein GMAC Brief #10
Relationship of preadmission measures to academic success in graduate management education	D. Powers F. Evans RB-78-11
Predicting career progress of MBAs seven years after graduation: Eleven-school sample	L. Crooks J. Campbell D. Rock (in press)
A prototype report for the GMAC Validity Study Service	D. Powers PR-77-13
Designing your validity study: A manual for the Graduate Management Admission Council Validity Study Service	D. Powers F. Evans
Final report of the results of the GMAC Validity Study Service pilot study	D. Powers F. Evans
Relationships of preadmission measures to academic success in graduate management education	D. Powers F. Evans RB-78-11

PR (Project Report)
RB (Research Bulletin)

Reports Available for GREB Research Projects Completed During the Last Five Years

Comparison of a Bayesian and a least squares method of educational prediction (1975)	R. Boldt RB-75-15
The Graduates: A report on the characteristics and plans of college seniors (1973) careers and curricula: A report on the activities and views of graduates a year after leaving college (1974) (Cooperatively sponsored by GREB, AAMC, and LSAC)	L. Baird GREB-70-4R
Factors in graduate student performance (1974)	R. Reilly RB-74-2 GREB-71-2P
A study of the relative effectiveness of a program of coaching/instruction for the GRE-Q for black, Chicano, and white GRE candidates (1977)	F. R. Evans GREB-71-5aP
Word associations of students at predominantly white and predominantly black colleges (1975)	J. Campbell L. Belcher RB-75-29 GREB-71-6P
Women, men and the doctorate (1974)	J. Centra GREB-71-10R
Development of measures for the study of creativity (1975)	N. Frederiksen W. C. Ward RB-75-18 GREB-72-2-P
The development and pilot testing of criterion rating scales (1976)	A. Carlson R. Reilly GREB-73-1P
Summary of data from the graduate programs and admissions manual, 2nd Edition (1975)	R. Altman M. Wallmark GREB-74-1R
An investigation of the feasibility of obtaining additional subscores on the GRE advanced psychology test (1976)	M. McPeek R. Altman GREB-74-4P
A study of the predictive validity of the tests of scientific thinking (1977)	W. C. Ward N. Frederiksen RB-77-6 GREB-74-6P

Construct validity of free-response and machine-scorable versions of a test of scientific thinking (1978)	W. C. Ward N. Frederiksen S. B. Carlson RB-78-15 GREB-74-8P
Development of a test of global vs. articulated thinking: The Figure Location Test (1978)	T. F. Donlon R. R. Reilly J. D. McKee GREB-74-9P
The program review practices of university departments (1977)	M. J. Clark GREB-75-5aR
How universities evaluate faculty performance: A survey of department chairpersons (1977)	J. A. Centra GREB-75-5bR
The validation of GRE scores as predictors of first-year performance in graduate study: report of the GRE cooperative validity studies project (1979)	K. M. Wilson GREB-75-8R
Implicit guessing strategies of GRE aptitude examinees classified by ethnic group and sex (in press)	L. W. Pike GREB-75-10P
A factor analytic study of the GRE Aptitude Test (1977)	D. Powers S. Swinton A. Carlson GREB-75-11P
Admissions test scores as predictors of career achievement in psychology (1978)	W. B. Schrader GREB-76-1R
Factors contributing to the career performance of recent Ph.D.'s (1979)	M. J. Clark J. A. Centra GREB-76-2R
The implications of test speededness (in press)	T. F. Donlon GREB-76-9aP (Report) GREB-76-9bR (Bibliography)
A longitudinal study of the role of cognitive styles in academic evolution during the college years	H. Witkin GREB-76-10R

The effect of nonstandard undergraduate assessment and reporting practices on graduate admission processes (1978)	J. E. Knapp I. B. Hamilton GREB-76-14R
A factor analytic investigation of seven experimental analytical item types (1978)	D. E. Powers GREB-77-1P
Development of an inventory of documented accomplishments: Phase II (1979)	L. Baird GREB-77-3R
The criterion problem: what measure of success in graduate education? (1979)	R. T. Hartnett W. W. Willingham GREB-77-4R
A factor analytic study of the restructured GRE Aptitude Test (in press)	D. E. Powers S. S. Swinton GREB-77-6P
An analysis of time related decrements and recoveries for GRE repeaters across ability and sex groups (in press)	D. A. Rock GRE-77-9R

RB (Research Bulletin)
GREB No. (number of report)

LSAC Research Projects Completed During the Last Five Years

Factor analysis of law school grades	R. Boldt LSAC-73-1
A study of the effects of moderator variables on the prediction of law school performance	F. Evans D. Rock LSAC-73-2
An exploratory study of the possibility of curvilinearity in law school validity study data	B. Pitcher LSAC-73-3
Extended study of the relationship of selected transcript information to law school performance	R. Reilly D. Powers LSAC-73-4
Effect of differences in college grading standards on the prediction of law school grades	W. Schrader B. Pitcher LSAC-73-5
Predicting law school grades of black American law students	W. Schrader B. Pitcher LSAC-73-6
The predictive effectiveness of several experimental item types and the operational item types in the Law School Admission Test in 1970-71	W. McPeek B. Pitcher A. Carlson LSAC-74-1
LSAT part-score validity study	B. Pitcher LSAC-74-2
Predicting law school grades for female law students	B. Pitcher LSAC-74-3
A report on the development of behaviorally-anchored rating scales for use in law school admission	R. Reilly LSAC-74-4
A study of the effectiveness of empirical item option weighting on the LSAT	R. Reilly LSAC-74-5
Relationships between characteristics of minority group students who entered law school in 1968, 1969, and 1970	W. Schrader LSAC-74-6
Extension of the study of the possibility of curvilinearity in law school validity study data	W. Schrader B. Pitcher LSAC-74-7

Prediction of law school grades for Mexican American and black American students	W. Schrader B. Pitcher LSAC-74-8
Male-female LSAT candidate study: 1969-73	L. Wightman LSAC-74-9
Test bias and the prediction of grades in law school	R. Linn LSAC-75-1
DeFunis revisited: A psychometric view	H. Breland LSAC-75-2
A further study of predicting law school grades for female law students	B. Pitcher LSAC-75-3
Preferential law school admissions and the equal protection clause: An analysis of the competing arguments	M. Redish LSAC-75-4
Relationships among law school predictors, law school performance, and Bar examination results	A. Carlson C. Werts LSAC-76-1
Predictor score regions with significant differences in predicted law school grades from subgroup regression equations	R. Linn B. Pitcher LSAC-76-2
The validity of two experimental item types: Issues and facts, and artificial language	B. Pitcher M. McPeek M. Binkley LSAC-76-3
Efficacy of undergraduate grade adjustment for improving the prediction of law school grades	R. Boldt LSAC-76-4
Law School Admission Test: Comparisons of Canadian candidates with non-Canadian candidates	L. Wightman LSAC-76-5
Subgroups validity study	B. Pitcher LSAC-76-6
Cumulative grade-point average percentile rank as a predictor of first-year average grades in law school	B. Pitcher LSAC-76-7
Summary of law school validity studies, 1948-1975	W. Schrader LSAC-76-8
Applications and admissions to ABA accredited law schools: An analysis of national data for the class entering in the fall of 1976	F. Evans LSAC-77-1

A survey of the relevance of legal training to law school graduates	L. Baird LSAC-77-2
Comparing predictions of law school performance for black, Chicano, and white law students	D. Powers LSAC-77-3
On the treatment of multiple scores for Law School Admission Test repeaters	R. Linn LSAC-77-4
The validity of the Law School Admission Test scores for repeaters	B. Pitcher LSAC-77-5
Evaluation of three methods for treating repeaters' scores	R. Boldt LSAC-77-6
The validity of the usage experimental item type	B. Pitcher W. McPeek M. Binkley LSAC-77-7
Use of empirical Bayes methods	D. Rubin LSAC-78-1
Defining competence in legal practice: Report of a national survey of solo and small firm practitioners	R. Powell A. Carlson RB-78-3
Defining competence in legal practice: The evaluation of lawyers in large firms and organizations	L. Baird A. Carlson R. Reilly R. Powell
Access to graduate and professional schools: An annotated bibliography	LSAC (Microfiche production)

Reports for Projects LSAC-72-1 through LSAC-74-9 appear in Reports of LSAC Sponsored Research: Volume II, 1970-1974; reports for Projects LSAC-75-1 through LSAC-77-7 appear in Volume III, 1975-1977.

RB (Research Bulletin)

TRUTH IN TESTING ACT OF 1979; THE EDUCATIONAL TESTING ACT OF 1979

WEDNESDAY, AUGUST 1, 1979

House of Representatives,
Subcommittee on Elementary, Secondary,
and Vocational Education,
Committee on Education and Labor,
Washington, D.C.

The subcommittee met, pursuant to notice, at 9:30 a.m., in room 2175, Rayburn House Office Building, Hon. Carl D. Perkins (chairman of the subcommittee) presiding.

Members present: Representatives Perkins, Weiss, Kildee, Williams, Kogovsek, and Goodling.

Staff present: John F. Jennings, counsel; Richard DiEugenio, minority legislative associate; Laurence A. Uzzell, senior research specialist; and Martin LaVor, minority senior legislative associate.

Mr. Weiss. The Subcommittee on Elementary, Secondary, and Vocational Education will come to order.

Mr. Perkins, who chairs both the subcommittee and the full Committee on Education and Labor, will be with us shortly.

In the meantime, Mr. Goodling and I will start the hearing.

We have a large number of witnesses scheduled for this morning, and I think that we will try to do the same thing we did yesterday which I think worked out fairly well.

We will ask all of the witnesses to join us at the witness table and then whenever one is finished with his statement, which hopefully will be limited to 5 or 6 minutes at the most, we will then address questions to either the entire panel or to an individual witness.

I think that that way we get the benefit of the clash of ideas as well as the lack of repetition of questions which would otherwise obviously take place.

Some of the witnesses may not have arrived yet, especially those who may have been flying in, but we will ask all of you now who are here to please join us.

Senator LaValle, I understand, has not yet arrived, has he?

Senator Halperin, if you will join us.

Senator Vann has not yet arrived.

Mr. McKeany of the American Federation of Teachers; Mr. Herndon of the National Education Association.

Mr. Kovacs I think will be accompanying Mr. Herndon.

Dr. Graham of the council on dental education, American Dental Association; Mr. Millard H. Ruud of the American Law Schools Association; Mr. Jon Haber, consultant on standardized testing, U.S. Student Association, accompanied by Andrew Burnett; Dr. Roger T. Lennon, assistant to the chairman in measurement, Harcourt Brace Jovanovich, Inc.

If any of you have prepared statements which have not yet been submitted to us the staff will accept them from you, and we will proceed.

We will wait until the entire New York delegation is here before we commence with them, and I gather Mr. McKeany is substituting a statement.

Mr. McKeany is not going to be testifying directly this morning. He is going to be submitting a statement.

If you will begin, Mr. Herndon, and then I will call on everyone else in order.

STATEMENT OF SENATOR KENNETH LaVALLE, NEW YORK STATE SENATE, ACCOMPANIED BY SENATOR DONALD HALPERIN, NEW YORK STATE SENATE; TERRY HERNDON, EXECUTIVE DIRECTOR, NATIONAL EDUCATION ASSOCIATION, ACCOMPANIED BY FRANK KOVACS, DIRECTOR OF RESEARCH; JAMES W. GRAHAM, COUNCIL ON DENTAL EDUCATION, AMERICAN DENTAL ASSOCIATION; MILLARD H. RUUD, EXECUTIVE DIRECTOR, ASSOCIATION OF AMERICAN LAW SCHOOLS; JON HABER, CONSULTANT ON STANDARDIZED TESTING, U.S. STUDENT ASSOCIATION, ACCOMPANIED BY ANDRE BURNETT, DIRECTOR, NATIONAL THIRD WORLD COALITION, USSA; ROGER T. LENNON, ASSISTANT TO THE CHAIRMAN IN MEASUREMENT, HARCOURT BRACE JOVANOVICH, INC.

STATEMENT OF TERRY HERNDON, EXECUTIVE DIRECTOR, NATIONAL EDUCATION ASSOCIATION, ACCOMPANIED BY FRANK KOVACS, DIRECTOR OF RESEARCH

Mr. HERNDON. Thank you, sir.

NEA is pleased to have the invitation to appear today to discuss the matter of truth in testing and particularly the Educational Testing Act of 1979.

The delegates at the most recent NEA convention which occurred this past month strongly urged a congressional investigation of the standardized testing industry.

They ask that you also look at the tax exempt status of some aspects of the testing business and most particularly that the Congress look at the need for Federal truth in testing legislation to guarantee access to all tests and statistical material held by such agencies which affect students, parents, and teachers, so we certainly commend you, Mr. Weiss and the Congress, for addressing this issue.

The teachers have questioned the impact of standardized testing for a long time. In many respects we look at tests as much as one might look at narcotics. They probably have legitimate uses but they are so widely abused today that we have great concern about them.

That is not a recent concern, because Oscar Buros, who has been one of the giants of that business for a long time and the editor of the Mental Measurement Yearbook, has over the past years stated many of our contemporary concerns and has done so directly for the testing industry.

Shortly before his death, when he addressed the American college testing program on the personnel involved with that program at the University of Iowa, he made several observations, among which are publishers of secured tests not commercially available have been very good about making their confidential test available to persons invited to review the test for the Mental Measurement Yearbook.

He goes on to point out the information available to permit an adequate assessment by them is unsatisfactory, even for their needs, and since the tests could only be examined in highly restrictive conditions, if at all, it was most important that additional detailed information be generally available to researchers and examiners and educators and others who would evaluate the tests for the purposes for which they are being used.

He suggested that if that detailed information were made available simultaneously with the publication of tests, not only would educators and testing specialists be in a better position to evaluate the test and to interpret the results more intelligently but the publishers would be compelled to construct better tests pointing out both their strengths and their weaknesses.

There have been roadblocks to that. The roadblocks are placed in the way of the test user, the educator, and the researcher in deciding just what these tests mean and how they ought to be used, so we encourage the Congress to proceed with this act, to make the information available so that we and our members, researchers, and evaluators of tests across the country can examine them and can draw rational conclusions as to how the use of standardized tests is in fact restricting the equality of educational opportunity for all people.

We supported the enactment of the statute that was recently passed by the New York State Legislature, and I note in closing here today that, even though many of the institutions who make up the testing industry endeavored to influence the legislature to reject that law.

Just this past week I appeared on the "Today" program in New York City, along with Dr. E. Belvin Williams, senior vice president of the Educational Testing Service. When asked why they had opposed the law, he indicated that they had changed their minds and that it was a good law, that it was a good idea for students to receive copies of the tests that they had taken.

I observe when one of the strongest proponents of standardized testing, that being the Educational Testing Service, and one of the strongest opponents, and the National Education Association, agree the law embodies sound principles, then the law probably embodies sound principles, and we urge you to proceed with its expeditious passage.

[The complete statement of Mr. Herndon follows:]

PREPARED STATEMENT OF TERRY HERNDON, EXECUTIVE DIRECTOR, NATIONAL EDUCATION ASSOCIATION

The National Education Association is pleased to have been invited to discuss the matter of truth in testing, most particularly the Educational Testing Act of 1979. The delegates at the 1979 NEA Annual Meeting, which occurred just this past month, strongly urge a congressional investigation of the standardized testing industry. They also ask that Congress examine the tax-exempt status of some aspects of the testing business, most particularly the need for federal truth-in-testing legislation to guarantee access to all tests and statistical material held by the agencies that affect students, parents, and teachers. Therefore, we certainly commend Mr.

Ted Weiss (N.Y.), Mrs. Shirley Chisholm (N.Y.), Mr. George Miller (Calif.), and the Congress for addressing this issue.

Teachers have questioned the impact of standardized testing for many years, and in many respects we look at tests as much as one might look at narcotics—they probably have legitimate uses, but they are so widely abused that we have great concern about it. That, however, is not a recent concern of teachers and professionals in the field. Oscar K. Buros, who was one of the giants of that business and the editor of the Mental Measurement Yearbook, had for over 47 years stated many of our contemporary concerns and often done so directly to the testing industry.

Shortly before his death in 1978, Buros made several cogent observations during an address to the American College Testing Program personnel at the University of Iowa.[1] Excerpts from that address are pertinent to our discussion today:

"Publishers of secure tests not commercially available have been very good about making their confidential tests available to persons invited to review the tests for the Mental Measurements Yearbook. I deeply appreciate the wholehearted cooperation which we have received from the American College Testing Program, College Entrance Examination Board, Educational Testing Service, and other organizations. Nevertheless, I would like to point out that the information available to permit an adequate assessment to be made of these secure tests is quite unsatisfactory. Although our reviewers generally receive some in-house material, which is not available to other educators and psychologists, even this material is inadequate.

"I would like to recommend that every secure test be accompanied by a manual describing the planning, construction, standardization, reliability, and validity of the test. Since the tests themselves can be examined only under highly restricted conditions, if at all, it becomes of greatest importance that detailed information be provided for secure tests. The important role which these tests play in influencing the lives of the examinees is much greater than the role of the commercially available tests. Yet far less information is available on these secure tests and, to make matters worse, much of the information which the publisher does possess is considered to be confidential—for in-house use only. If detailed information were made available simultaneously with the publication of tests, not only would educators and testing specialists be in a better position to evaluate the tests and to interpret the results more intelligently, but the publishers would be impelled to construct better tests, pointing out both their strengths and weaknesses."

In another part of the same presentation, Buros identifies another problem that the testing industry has not corrected:

"Let me mention two important programs as examples of test which provide far too little information. These programs are probably no worse than many of the other programs of secure tests. I refer to the College-Level Examination Program and the ACT Proficiency Examination Program, both of which provide examinations to permit students to receive college credit without taking the relevant college courses. The CLEP program, initiated thirteen years ago, is one of the most significant developments in testing in recent years. It has received extraordinarily wide publicity—it is regularly advertised on television, yet there is relatively little research data on the validity of these tests. Roadblocks have been placed between the test user and what the scores mean. Only confidential house papers present raw score distributions. Such information should be routinely presented for every achievement test. The corresponding ACT PEP program is especially poor in the information it provides. I find it discouraging that two outstanding not-for-profit testing organizations have not provided test users with the hard data to permit them to assess more adequately the usefulness of the tests for the assignment of college credit.

"I would like to repeat a statement which I made forty-two years ago: Today it is practically impossible for a competent test technician or test consumer to make a thorough appraisal of the construction, validation, and use of most standardized tests being published because of the limited amount of trustworthy information supplied by test publishers and authors . . . If testing is to be of maximum value to schools, test authors and publishers must give more adequate information * * * It would be advantageous * * * if test publishers would construct only one-fourth to one-half as many tests * * * and use the time saved for presenting the detailed information needed by test consumers.

"Unfortunately, although some progress has been made, my 1935 complaint is equally applicable today to the majority of existing tests—and especially so for secure tests."

---

[1] Buros, Oscar K. "Fifty Years in Testing: Some Reminiscences, Criticisms, and Suggestions." *Educational Researcher* 6: 9-15; July-August 1977.

The NEA believes that the testing industry will not move to correct these inappropriate practices without federal legislation. We would encourage the Congress to proceed with H.R. 4949 and make the information available so that we and our members, researchers, and evaluators of tests across the country can examine them, and can draw rational conclusions on how the use of standardized tests is restricting the equality of educational opportunity for all people.

NEA supported the enactment of the statute passed by the New York State legislature; and even though many of the institutions who make up the testing industry endeavored to influence the legislature to reject that law, just this past week I was a guest on NBC's Today program in New York City along with Dr. E. Belvin Williams, senior vice-president of the Educational Testing Service. When asked why ETS has opposed the law, Dr. Williams indicated that ETS had since changed its mind, that it was a "good law," and that "students should see their tests." So, when one of the strongest proponents of standardized testing, that being the Educational Testing Service, and one of the strongest opponents, that being the National Education Association, agree that the law embodies sound principles, then the law probably embodies sound principles.

NEA urges the Congress to proceed and expedite the passage of this legislation.

Mr. WEISS. Thank you very much.

I understand Senator LaValle has now arrived.

Senator, we welcome you and I just want to give a word of introduction to both Senator LaValle and Senator Halperin.

Senator Halperin is a personal as well as a political friend of many years standing.

Senator LaValle is one who I know by reputation, which is a very esteemable one. They represent both the minority and the majority in the New York State Senate and have been successful after a couple of years, I guess, of efforts in having some major legislation adopted by the State, the Senate and the State Assembly and signed into law by Governor Carey just a few weeks ago.

We will ask Senator LaValle to proceed first with his statement and ask Senator Halperin to make his statement, but the questions will be delayed until the entire panel has completed its testimony.

STATEMENT OF SENATOR KENNETH LaVALLE, NEW YORK STATE SENATE

Mr. LAVALLE. Thank you, Congressman Weiss.

My name is Ken LaValle and I represent the First State Senatorial District in New York.

I am pleased to appear before you today as you probe the important issues surrounding testing, issues that greatly affect millions of postsecondary students throughout our Nation each year.

As chairman of the New York Senate Higher Education Committee, I held a joint hearing with the Assembly Higher Education Committee in May on New York's Truth-in-Testing legislation.

As the prime senate sponsor of that bill, I am very pleased to say that our hearing was only the first successful step in the life of that bill. We were able, with the help of several testing experts, as well as the support of parents, teachers, and students, to get our bill passed in both houses and eventually signed by the Governor.

Before I trace the history of what proved to be one of the most controversial issues of our 1979 session, let me say that I sincerely hope that truth-in-testing at the Federal level meets with similar success.

The facts are clear. This legislation is necessary and the timing is critical. Each student denied access to his or her exam is unfairly and irreparably injured.

Each unscrutinized test potentially destroys the future of thousands of students. These are the lives of our youth, our greatest natural resource.

I urge you to act now to prevent further harm.

I became interested in this issue really as an educator before I started my career in government, and one of the things that I think, having some knowledge and some experience in the testing area, that I found that there was a great importance placed on these tests and that many of us as educators were taught that often these tests had a very low statistical validity.

My interest continued and later, when I was executive director of the State's senate education committee, I was able to view the impact of this issue from a statewide perspective.

Finally, as State senator, I introduced the first truth-in-testing legislation in 1978. Although the bill remained in committee at the close of the 1978 session, it served to raise the issue and enlighten many legislators regarding the problems surrounding standardized admissions tests.

Off session hours were then spent in research which involved consultations with experts in the field. The bill I introduced this year was the result of this research and it reflected the advice we received.

I have to compliment the people on my staff, along with the people such as Senator Halperin and his staff, for the research and the field work that they did on this legislation.

The hearing we held lasted 6 hours and produced testimony which was overwhelmingly in support of the desire and need for this legislation.

Most of the opposition offered was based on fear, fear instilled by unsupportable threats of dire circumstances made by the major testing agencies.

As previously stated, the support was overwhelming and the testimony was replete with questions about:

The sufficiency of publicly available data for independent evaluations of construction, validity, or use of the tests;

The quality of the items appearing on the tests and the possible existence of built-in bias;

The many reports of known scoring errors, leaving great doubt as to the number of scoring errors left undiscovered, and, perhaps most compelling, was testimony concerning the issue of fairness.

Ought not the burden be on those denying access to prove the need to continue in secrecy?

The answer must be "Yes." I am here today to testify that that burden has not been met, nor will it be. I have heard all their arguments. I assure you that when you have heard them you will agree with me.

Defeat of our bill in New York was urged for the following reasons:

First, the technique of equating would be destroyed;

Second, the cost of compliance would cause student fees to rise as much as 50 percent and, finally, rather than comply, test agencies would boycott New York and as a result New York students would be unduly harmed.

Despite intensive lobbying, done of course with student fees, these arguments convinced neither the majority of the senate or assembly nor the Governor, and I must indicate that the lobbying on this legislation was probably some of the toughest that I have encountered, and some of you may have seen a recent piece by Jay Anderson on the lobbying by some of the test-making companies on this legislation, and I am sure that you will encounter the same kind of intense lobbying here on the Hill.

Current knowledge among testing experts disposed of all equating arguments. Use of methods already employed by major testing agencies such as the practice of simply not scoring equating questions or use of proven new techniques would permit equating to continue undisturbed. No attempt was ever offered to refute these facts.

I might just emphasize that there are—and some of the experts have indicated that there are—equating techniques that have not been used, and it would seem there is almost a laziness on the part of the testing people not to go in and use those techniques.

Similarly, the cost issue was unconvincing, particularly in light of undisputed information regarding the small percentage of student fees used for test development and the very large surplus earned by the major companies on these tests.

As you all know, the testing companies, as chartered in our State, New York, are nonprofit, tax-exempt corporations.

For example, ETS documents indicate that only 5 percent of the student fee for the SAT and only about 7 percent of the student fee for the GRE is used for test development, the only aspect of the testing process affected by our bill. In addition, their documents show earned surplus in excess of 22 on the SAT, a figure estimated to be similar for other testing programs.

Information that I requested and received from ETS following our hearing stated that the bill would increase SAT development by $1,092,000. Even if ETS's figure is accepted, despite our conflicting data, it still amounts to less than one-third of the surplus taken by ETS and College Board on the SAT. Between 1974-75 and 1976-77 the earned surplus ranged from $3 million to $3,541,000.

Upon learning of the extremely profitable nature of this industry, many legislators asked both me and our staff, well, if we have this, if they have this surplus, why don't we legislate a decrease in the cost of these examples?

The third argument in opposition, that of possible isolation of New York, was viewed as merely a desperate threat since such a course of action would be an unsound business decision.

However, as we now know, the American Association of Medical Colleges and the Dental Association have reacted to the adoption of this legislation by stating they would no longer sponsor the MCAT's or DAT's in New York. This attempt to avoid the disclosure legislation of one State points out the need for Federal legislation in this area.

I would like to take this opportunity to say that neither of these professional associations participated in the debate generated by this bill.

Although the spokesperson for Associated Medical Colleges of New York mentioned affiliation with the American Association of

Medical Colleges at our hearing, we have, to this date, received no communication from the Dental Association and we did not hear from the AACM for nearly a week following their announcement.

Reasons for their noninvolvement are not clear, and I would hope an explanation will be offered at this hearing. If they persist in this conduct, however, I intend to use my subpena powers as chairman of the Senate Higher Education Committee to compel them to substantiate their claims.

For example, they claimed there are finite numbers of quality questions which may be used on these tests. We disposed of this argument when it was raised concerning other tests covered by the legislation, and I believe we can similarly prove it is without merit in the case of the MCAT's and the DAT's.

Can we be expected to believe that science, the most dynamic field known to man, cannot yield enough questions for two tests each year? We have been through the manual, and I am going to read some of the examples, but the staff has looked at the manual that is produced by MCAT and we looked at some of the questions, and I think the committee will find that many of those questions can be reproduced without any problem.

Mr. **Weiss**. Without objection, Senator, whatever additional material you would like to have submitted for the record will be entered.

Mr. **LaValle**. If the AAMC and Dental Association are threatening to boycott New York merely out of arrogance, then I believe they must state it as such. The fact is, they can comply without compromising their exams. Passage of Federal legislation will, of course, insure compliance, and I emplore this committee and Congress to act swiftly.

At this point I would like to comment on Congressman Weiss' bill which has been drafted, I believe, with the aid of some of the experts in our bill.

I am particularly interested in section 7, entitled "Testing Costs and Fees to Students."

The information that would be disclosed pursuant to this section would answer many of the questions raised by the conflicting memos circulated for the New York version of truth-in-testing.

I requested many of the figures which would be required for disclosure but was unable to receive them. A constant accounting by the testing agencies would prohibit retaliatory actions for ostensibly meritorious reasons.

To the degree that these testing agencies have such a tremendous impact on the lives of millions each year, I believe this kind of monitoring is desirable and, in fact, necessary.

The testing industry has shrouded itself for years in what I have called a "mantle of secrecy," and this provision as well as others in the bill would certainly help pierce that veil.

I am also very interested in section 8 which would amend the Higher Education Act of 1965 to require institutions to explain their use of test scores in the admissions process.

Mr. **Weiss**. That is the provision that we have deleted from the final half of the bill, and we no longer include it.

Mr. **LaValle**. All right; we will skip down and just say in conclusion——

Mr. WEISS. Not everybody was unanimous in his appreciation as you are.

Mr. LAVALLE. In conclusion, let me urge you to listen with a critical ear to the claims of the test agencies before you. Independent experts have seen the need for this legislation for years; students have felt the need and are calling for your aid.

If they are to be judged by a single instrument, if their futures are to be determined by a few hours of highly pressurized effort, let that test be the best it can be.

Passage of this legislation is the most positive step you can take toward that goal.

I commend your decision to hold this hearing on this most important matter, and thank you for inviting me to testify.

[The prepared statement of Kenneth LaValle follows:]

PREPARED STATEMENT OF SENATOR KENNETH LAVALLE, NEW YORK STATE SENATE

My name is Ken LaValle and I represent the first State senatorial district in New York. I am pleased to appear before you today as you probe the important issues surrounding testing—issues that greatly affect millions of postsecondary students throughout our Nation each year.

As chairman of the New York Senate Higher Education Committee, I held a joint hearing with the assembly Higher Education Committee in May on New York's truth-in-testing legislation. As the prime senate sponsor of that bill, I am very pleased to say that our hearing was only the first successful step in the life of that bill. We were able, with the help of several testing experts, as well as the support of parents, teachers, and students, to get our bill passed in both houses and eventually signed by the Governor.

Before I trace the history of what proved to be one of the most controversial issues of our 1979 session, let me say that I sincerely hope that truth-in-testing at the Federal level meets with similar success. The facts are clear—this legislation is necessary and the timing is critical. Each student denied access to his or her exam is unfairly and irreparably injured. Each unscrutinized test potentially destroys the future of thousands of students. These are the lives or our youth—our greatest natural resource. I urge you to act now to prevent further harm.

I became interested in this issue years ago as both an educator and a parent. The field of measurement is not unfamiliar to me, and I watched with great concern the proliferation of test administrations and the increasing importance assigned to tests I had been taught were often of low statistical validity.

My interest continued and later, when I was executive director of the State's Senate Education Committtee, I was able to view the impact of this issue from a statewide perspective. Finally, as State senator I introduced the first truth-in-testing legislation in 1978. Although the bill remained in committee at the close of the 1978 session, it served to raise the issue and enlighten many legislators regarding the problems surrounding standardized admissions tests. Off session hours were then spent in research which involved consultations with experts in the field such as Dr. Vito Perrone, president of the national consortium on testing and Dr. Lewis Pike, senior associate at the National Institute of Education and former ETS employee for 14 years. The bill I introduced this year was the result of this research and it reflected the advice we received.

The hearing we held lasted six hours and produced testimony which was overwhelmingly in support of the desire and need for this legislation. Most of the opposition offered was based on fear—fear instilled by unsupportable threats of dire circumstances made by the major testing agencies. As previously stated, the support was overwhelming. The testimony was replete with questions about:

The sufficiency of publicly available data for independent evaluations of construction, validity or use of the tests;

The quality of the items appearing on the tests and the possible existence of built-in-bias;

The many reports of known scoring errors—leaving great doubt as to the number of scoring errors left undiscovered; and

Perhaps most compelling, was testimony concerning the issue of fairness. Ought not the burden be on those denying access to prove the need to continue in secrecy?

The answer must be yes. I am here today to testify that that burden has not been met—nor will it be. I have heard all their arguments. I assure you that when you have heard them you will agree with me.

Defeat of our bill in New York was urged for the following reasons: The technique of equating would be destroyed; the cost of compliance would cause student fees to rise as much as 50 percent; and finally, rather than comply, test agencies would boycott New York and as a result, New York students would be unduly harmed.

Despite intensive lobbying—done of course with student fees—these arguments convinced neither the majority of the senate or assembly nor the Governor.

Current knowledge among testing experts disposed of all equating arguments. Use of methods already employed by major testing agencies such as the practice of simply not scoring equating questions or use of proven new techniques would permit equating to continue undistured. No attempt was ever offered to refute these facts.

Similarly, the cost issue was unconvincing, particularly in light of undisputed information regarding the small percentage of student fees used for test development and the very large surplus earned by the major companies on these tests.

For example, ETS documents indicate that only 5 percent of the student fee for the SAT and only about 7 percent of the student fee for the GRE is used for test development—the only aspect of the testing process affected by our bill. In addition, their documents show earned surplus in excess of 22 percent on the SAT—a figure estimated to be similar for other testing programs. Information that I requested and received from ETS following our hearing stated that the bill would increase SAT development by $1,092,000. Even if ETS's figure is accepted, despite our conflicting data, if still amounts to less than one-third of the surplus taken by ETS and college board on the SAT. Between 1974-75 and 1976-77 the earned surplus ranged from $3,000,000 to $3,541,000.

Upon learning of the extremely profitable nature of this industry, many legislators asked: Why don't we legislate a decrease in the cost of these exams?

The third argument in opposition—that of possible isolation of New York—was viewed as merely a desperate threat since such a course of action would be an unsound business decision.

However as we now know, the American Association of Medical Colleges and the Dental Association have reacted to the adoption of this legislation by stating they would no longer sponsor the MCAT's or DAT's in New York. This attempt to avoid the disclosure legislation of one State points out the need for Federal legislation in this area.

I would like to take this opportunity to say that neither of these professional associations participated in the debate generated by this bill. Although the spokesperson for Associated Medical Colleges of New York mentioned affiliation with the American Association of Medical Colleges at our hearing, we have, to this date, received no communication from the Dental Association and we did not hear from the AAMC for nearly a week following their announcemnt. Reasons for their noninvolvement are not clear and I would hope an explanation will be offered at this hearing. If they persist in this conduct, however, I intend to use my subpena powers as chairman of the Senate Higher Education Committee to compel them to substantiate their claims. For example, they claimed there are finite numbers of quality questions which may be used on these tests. We disposed of this argument when it was raised concerning other tests covered by the legislation, and I believe we can similarly prove it is without merit in the case of the MCAT's and DAT's. Can we be expected to believe that science—the most dynamic field known to man—cannot yield enough questions for two tests each year? Let me cite examples of questions appearing in "The New MCAT Student Manual":

These are examples of questions so unique that they defy reproduction? Obviously not.

If the AAMC and Dental Association are threatening to boycott New York merely out of arrogance, then I believe they must state it as such. The fact is, they can comply without compromising their exams. Passage of Federal legislation will of course insure compliance and I implore you to act swiftly.

At this point I would like to comment on Congressman Weiss's bill which has been drafted, I believe, with the aid of some of the experts involved in our bill. I am particularly interested in section 7 entitled, "Testing Costs and Fees to Students." The information that would be disclosed pursuant to this section would answer many of the questions raised by the conflicting memos circulated for the New York version of truth-in-testing. I requested many of the figures which would be required for disclosure but was unable to receive them. A constant accounting by the testing agencies would prohibit retaliatory actions for ostensibly meritorious reasons. To the degree that these testing agencies have such a tremendous impact on the lives of millions each year, I believe this kind of monitoring is desirable and in fact, necessary. The testing industry has shrouded itself for years in what I have called a "mantle of secrecy" and this provision as well as others in the bill would certainly help pierce that veil.

I am also very interested in section 8 which would amend the Higher Education Act of 1965 to require institutions to explain their use of test scores in the admissions process. ETS was quick to declare that they do not advise use of test scores as the sole criterion for admission—not a surprising caution since they never denied that academic grades are by themselves a better predictor of future success in school.

Despite this advice, I have learned that admissions officers are using test scores to establish arbitrary cut-off points for admissions decisions. While the reasons for this are many, it is significant in that it portrays the attitude of the general public. The major test companies have become so complex that their product is often granted unwarranted value, regardless of their statistical validity. This provision, requiring disclosure of the admissions procedure followed by each institution will further assist students in their decision process and will hopefully encourage the higher education community to re-evaluate their current practices.

In conclusion, let me urge you to listen with a critical ear to the claims of the test agencies before you. Independent experts have seen the need for this legislation for years: students have felt the need and are calling for your aid. If they are to be judged by a single instrument—if their futures are to be determined by a few hours of highly pressurized effort, let that test be the best it can be. Passage of this legislation is the most positive step you can take toward that goal. I commend your decision to hold this hearing on this most important matter, and thank you for inviting me to testify.

Mr. **Weiss**. Senator, thank you very much for your testimony.

Before I ask Senator Halperin to give us his statement, let me explain for those of you who may be new to our bells and buzzers and lights on the clock that the House is now in session.

We are also subject to being summoned to some floor action for a vote, or whatever.

When that happens the bells, the buzzers go off, and we have 15 minutes from the time that the first bell sounds to get down there and so we may have to from time to time recess the proceedings, but we will complete the session and hopefully the disturbances or delays will be minimal.

With that, Senator Halperin, please proceed.

STATEMENT OF SENATOR DONALD HALPERIN, NEW YORK STATE SENATE

Mr. **Halperin**. Mr. Chairman and honorable members of the subcommittee, I appreciate the opportunity to testify here today in support of Federal legislation which would establish disclosure requirements for the developers of standardized admission tests for university and postgraduate schools, similar to those recently adopted by New York State.

As a member of the New York State Senate Standing Committee on Higher Education, and a former ranking Democratic member of that committee and a cosponsor of the New York legislation, I was deeply involved in its passage.

While the New York law and the proposals pending before you differ in a number of ways, my purpose in testifying is to convince you of the advisability of this approach. I have full confidence that should you decide to adopt legislation the technicalities can be worked out.

I would add that since this statement was prepared, I had an opportunity to review Congressman Weiss' bill and I do believe that it effectively implements the concept.

The importance of these tests to millions of Americans is inescapable. Many schools rely heavily upon them in making admissions decisions, and a poor grade can mean nonconsideration.

The proposed legislation should be supported for three basic reasons:

The public and outside experts should have the opportunity to evaluate the effectiveness of these tests in predicting scholastic success and accomplishment potential;

Test takers should have the opportunity to identify administrative and clerical errors, and all test takers should have an equal opportunity to evaluate their test-taking ability and to understand the nature of the test.

Let me now deal with the issue of validity, a basic question since it goes to the very purpose of the test.

I have very strong reason to be a proponent of standardized tests since they always have indicated a great aptitude on my behalf than did my grades. But there are many examples to the contrary.

One such example is a young man who worked in my law office as a clerk. When he took the SAT he got such low grades that he was told a monkey could have done as well. Yet he was valedictorian of his class in high school.

One good college was willing to take a chance and he was accepted. He did fantastically well. He made dean's list and was considered for a number of very fine fellowships.

Then it came time to go to law school, so he took the LSAT and once again ranked as high as a Rhesus monkey.

This time he wasn't as lucky and no law school was willing to consider him even with his grades. He then took the job at my law firm and took the test again.

This time his average went up a few points, according to him, merely a matter of luck. After a renewed search he found a school willing to accept him.

I recently called to find out how he was doing in law school. Not quite as well as college; not a valedictorian, but he successfully completed his first semester and felt he was doing much better in his second semester.

It is absolutely unfair to him and others who have similar problems to be cut off from any knowledge as to how they are being evaluated and why they are unsuccessful. But that is happening to thousands of students.

This type of test-taking problem is but one example of poor predictability. What about geographic and ethnic factors which might not be properly accounted for? Are we to rely solely on those whose business it is to sell the validity of their process?

I think that is poor public policy. The process must be open to outside evaluation.

I do not think it necessary to dwell upon the next issue; administrative and clerical error. Suffice it to say we have all had our own unfortunate experiences with human and computer error. At least we are usually given the benefit of explanations and the opportunity to fight for our just due.

Examples do exist of error in scoring. These were discovered by fortuitous happenstance. Who knows how many lives were unfairly altered by undetected and unprovable error?

The third and final reason I offer for support of the proposed legislation is equalized access to test questions and techniques by all test takers. In this regard I was a beneficiary of the existing system. I was one of those whose family was able to afford a course which helped to raise my grades from the PSAT to the SAT by 200 points in English and by over 100 points in math.

A well known and highly organized tutoring course that I was able to participate in had people taking the test year after year, extracting questions which, as the test people themselves say, are used year after year.

A whole list of questions were accumulated and our course entailed going over those questions and knowing how to answer them. So when I saw many of them for the first time I was no stranger to them.

To accept the argument raised in opposition to the proposed legislation that public disclosure of the questions will undermine the validity of the test, is to accept a myth. These questions are already to a large degree public, but they are available only to those lucky enough to know about tutoring course, able to afford these courses, and geographically accessible to them.

It has also been argued that certain questions are necessary to equate the results of groups who take the test at different times. This issue is easily dealt with by excluding from the legislation's coverage those equating questions and then not using such questions for the purpose of scoring.

I am not quick to encourage the entry of the Federal Government into areas I feel should be left to State regulations but I feel the time has clearly arrived for Congress to take action in this area.

One reason is that New York is now being blackmailed by the powerful Association of American Medical Colleges with the threat that they will not conduct their tests in our State because of the new law.

This threat would dissolve with the adoption of Federal legislation, but more importantly these tests are used to evaluate and consider students from different parts of the country and, therefore, should have national standards.

Also, access to tests must become a national policy so that their integrity is maximized and so that those using these instruments of predictability will have to do so with full knowledge of their strengths and their weaknesses.

Thank you, Mr. Chairman.

Mr. WEISS. Thank you very much, Senator.

Dr. Graham?

STATEMENT OF JAMES W. GRAHAM, ASSISTANT SECRETARY, COUNCIL ON DENTAL EDUCATION, AMERICAN DENTAL ASSOCIATION

Dr. GRAHAM. Thank you, Congressman Weiss.

Mr. Chairman and members of the subcommittee, my name is Dr. James Graham.

I am an assistant secretary of the American Dental Association's Council on Dental Education.

The American Dental Association conducts two national testing programs.

One, the dental admission testing program which is an entrance examination required by all the 60 dental schools in the United States, and the other, the council of national boards examinations, a proficiency examination required by 48 of the 50 States as a partial fulfillment for State licensure.

I would like to talk about each of the programs separately.

The dental admission testing program or DAT was begun in 1946 for the express purpose of reducing attrition taking place in dental schools.

During the 1930's and 1940's, before the test was instituted, the national attrition rate in dental schools averaged between 20 and 25 percent. After the DAT was accepted as a requirement for entrance for all dental schools in 1950, and additional requirements were introduced, this rate was reduced to about 7 percent.

If one were to convert this to present-day statistics, since there are now slightly over 6,000 first-year dental students, this attrition rate reduction would be from 1,200 to 420. With the average cost of educating a dental student now being about $19,000 per year, the American Dental Association believes the dental admission testing program has more than achieved its goals.

The dental admission testing program is in many ways a very unique admission testing program. First, the program is conducted entirely in-house by the American Dental Association's Council on Dental Education.

It maintains no relations or subcontracts with any of the large testing corporations or test consulting firms nor has it ever maintained any such relations or subcontracts.

The dental admission testing program is the only postsecondary entrance testing program conducted by a professional organization like the American Dental Association.

It does this in order to maintain better control over the quality of its program at what it feels is a more reasonable price. Since it began in 1950 the program has increased its fee once and that was by only $5 to its present fee of $20. The committee should understand that compliance with this bill would substantially raise this cost to the students.

A second unique feature of the dental admission testing program is its manual dexterity tests. As you know, one of the absolutely necessary and basic skills required of all dental students is the ability called fine manual dexterity.

In order to graduate from dental school, the dental student must have the proficiencies to practice what is commonly called general dentistry. General dentistry, like the medical specialty of surgery, as opposed to the practice of most other professions, requires a high degree of manual technique proficiency.

Not everyone has the fine manual dexterity to master the technical skills to practice dentistry. The DAT has developed nonverbal, perceptual, and space relation tests which over the years have demonstrated proven ability to measure and evaluate the applicant's fine manual dexterity.

The formal of these tests is nonverbal in nature. It is a series of drawings of two and three dimensional objects which the test subject must relate to. The problem is that there are only a small number of ways to represent a box or triangle on a drawing, which severely limits the number of test forms that can be produced.

This is one of the primary reasons for our testimony today. If the dental admission testing program was required to disclose the actual tests after each administration, as the legislation now proposes, in a year or two these kinds of manual dexterity tests would have to be discontinued.

I would like to make several further points with regard to the issue of disclosure. The DAT is the only entrance examination program that contains tests that measure abilities other than academic. If the dental profession lost this instrument it would suffer irreparable harm, since this is the best proven criteria that can be used.

A diminution in quality of the other portions of the DAT, for example, its science tests or its reading comprehension tests could be compensated for in part by other criteria, like grades in college.

Without the manual component of the DAT, however, dental schools would have no criteria to discriminate ability to perform fine manual dexterity and as consequence it can be predicted that the attrition rate in dental schools would again rise to where it was before the DAT was introduced.

The American Dental Association asks you to give this point very special consideration in your deliberations.

It should also be made clear that the American Dental Association does not disagree with the objectives which the disclosure clause or the other provisions the proposed legislation is trying to accomplish. To the contrary, it fully endorses these objectives and has in fact taken steps which it feels are beyond those taken by most other testing programs to accomplish these objectives.

In fact, the American Dental Association feels that these are merely the objectives of testing which have for many years been clearly promulgated by organizations like the American Pyschological Association and the American Educational Research Association.

I will enumerate a few of the measures it has taken to accomplish these objectives:

First, it has released an example copy of an exact form of the dental admission tests which it has entitled, "Dental Admission Test Preparation Materials." These materials are distributed free of charge to all applicants who register for the DAT.

It has done this to neutralize the error measurement that has been introduced into the test scores by test preparation course, since all potential candidates, especially those whose economic circumstances, backgrounds or geographic location in the country, do not have identical access to preparation courses.

The ADA also feels that when a candidate studies for a test, he or she ought to be able to study the material which is actually contained in the test at the difficulty level it will be tested. This is especially important for the nonverbal, perceptual and space relation tests contained in the program.

The American Dental Association also believes that it is assisting the dental schools to more accurately judge the ability of the candidate by eliminating the above types of bias that might be present in their admission decisions.

Second, annually, the American Dental Association conducts a complete statistical analysis of its tests along with validity studies which are participated in by all the dental schools in the United States.

The statistical analysis contains a complete item-by-item analysis of the test to determine not only how reliable the tests were as a whole, but also how each specific item in the test performed. The analysis also contains additional studies to determine the effect of repeating the test more than one time.

These studies have been carried on since the inception of the program and are fully published and distributed free of charge to anyone who has requested copies of them. These studies along with other useful admission information are bound together under the title "Handbook for Pre-Dental Advisers" and annually distributed free of charge to all pre-dental advisers at all undergraduate post-secondary institutions across the country.

In this way, the American Dental Association feels that it has fully informed and made freely available all information that is necessary for an applicant or his adviser to judge the quality or content of its dental admission testing program. A copy of the Dental Admission Test Preparation Materials and other studies just mentioned have been submitted to the subcommittee for examination.

At this point I would like to briefly address our concerns as they relate to the national board examinations. The Council on National Board Examinations was established in 1928 as a standing committee of the American Dental Association for the purpose of providing and conducting written examinations for use at the discretion of State boards of dentistry in licensing dentists.

In 1961, the charge of the council was expanded to include providing and conducting written examinations for use at the discretion of State boards of dentistry in licensing dental hygienists. Currently, over 12,000 candidates for dental licenses and almost 6,000 candidates for dental hygiene licenses participate in National Board programs each year.

The dental examination battery consists of seven 100-item examinations, while the dental hygiene examination is a single examination consisting of 350 items. In the interest of serving the public, State boards of dentistry, candidates for licensure and the profession, examinations are administered three times at testing centers throughout the United States.

To insure examinations of high quality that are fair and just to participants, two or three new editions of each dental examination and one or two editions of the dental hygiene examination are developed each year.

During any year, at least four different dental examinations are released with answers and statistical analysis of test items to licensure candidates, dental school faculty members and interested individuals. Dental hygiene examinations are also released regularly, but with less frequency due to the length of examinations.

The American Dental Association supports in principle the disclosure provisions contained in the legislation. The American Dental Association believes that, through its current policy of releasing National Board examinations, it fully responds to the extent of the disclosure clause.

Our concern relates to having undue restrictions placed on its examination program that would jeopardize the quality of examinations. Publishing all test items after their use would seriously deplete the test item files and cause examinations to become less reliable and less valid.

The American Dental Association also urges the rejection of proposals which would stipulate that examinations used for entry into a profession must be scored without relation to the distribution of scores produced by the examination.

National Board examinations are scored using a distribution-referenced scoring system. This scoring system allows for consistency among scores and equivalency of scores from various editions of the examinations.

Until criteria-based scoring procedures are refined to provide identical interpretation of scores from testing date to testing date, a distribution-referenced scoring system is the fairest and most reliable way of measuring the cognitive skills necessary to practice dentistry or dental hygiene.

In summary, the American Dental Association does not believe that legislation such as H.R. 4949 is necessary for those programs which I have described.

If Congress should decide that some type of limited legislation is desirable, then the disclosure provision of sections 5(a)(1)(a) and 5(a)(2)(a) should be deleted or the exemption in section 5(b) should be extended to cover relatively small programs, such as the DAT which are adminstered to less than 10,000 subjects per testing period.

With regard to H.R. 3564, the American Dental Association likewise feels that this legislation is unnecessary as applied to the association's programs. If the legislation is to be enacted, it feels that section 6(b)(4) should be deleted in that it might be extended to include disclosure and that section 6(c) should be deleted since the theory of criterion reference tests has not been sufficiently developed to warrant its indiscriminate imposition on all proficiency testing programs.

Mr. Chairman, this concludes my remarks.

I would be happy to attempt to respond to any questions which the subcommittee may have.

Mr. WEISS. Thank you, Dr. Graham.

Mr. Millard Ruud?

STATEMENT OF MILLARD H. RUUD, EXECUTIVE DIRECTOR, ASSOCIATION OF AMERICAN LAW SCHOOLS

Mr. RUUD. Thank you, Mr. Chairman.

I am Millard H. Ruud, executive director of the Association of American Law Schools and for 30 years a member of the law faculty of the University of Texas.

While a member of that faculty, I served for about 15 years as the chairperson of the school's admissions committee and on various committees of the Law School Admission Council, the national organization that owns the Law School Admission Test and manages the law school admission program.

I served as its chairman from 1966 to 1969. Today, I appear as the representative of the Association of American Law Schools, AALS.

The Association of American Law Schools was founded in 1900 and is composed of 136 member law schools, all of which are approved by the American Bar Association. The association is recognized as a national accrediting body by the Council on Postsecondary Accreditation, a nonprofit, nongovernmental association. As stated in its bylaws, the purpose of our association is the "improvement of the legal profession through legal education."

I think, Mr. Chairman and members of the committee, that our association is the only one appearing before you that represents the admission committee users of the test, and so I reflect that perspective and the perspective of the concern of our schools for the use of the test and their concerns for their applicants and potential applicants.

I should, to help you understand the dynamics of the use of standardized admissions tests of the Law School Admission Council, I should emphasize that the Law School Admission Council is a separate entity incorporated under the education laws of New York, governed separately by representatives, by persons elected by representatives of the schools.

It conducts it services largely under a contract with the Educational Testing Services.

The operating relationships and the institutional relationships that these two organizations have are excellent.

While we are not unmindful or unsympathetic to the substantial problems that these proposals have for the Law School Admission Council, the testimony often is offered from the perspective of our member law schools.

I should emphasize that we have no economic interest in the LSAT, no economic interest in the Law School Admission Council, or any of its services.

My review of the two bills before this committee has disclosed to me three important themes or purposes, all of which are consistent with the interest of our association.

Let me turn briefly to these themes, as I see them.

One, a very important purpose of the bill enacted in California, the bill enacted in New York, and in the two bills before you is that of providing potential test takers with as much information about the admission test as possible, so as to place all who take the test on equal footing.

The special role that the admission test score plays is one item in making the admissions decision that of providing an impartial and nationally uniform basis for judging applicants.

None of the other information before the admissions committees of our law schools have a common and impartial basis.

Now, for the admissions test scores to truly have an impartial and common basis, it is essential that those who take the test be placed on as narrowly a common basis as possible.

Much of section 6 of H.R. 3564 and some of section 3 of H.R. 4949 are designed to serve that purpose.

In my written testimony I described in detail the efforts that are made, starting with publication by LSAC of 500,000 copies of this document, which contains a form of the test, correct answers, a description with respect to each type of question, of why a given answer is the correct one.

Mr. Weiss. Mr. Ruud, without objection, the entire written statement that you have will be entered into the record and, of course, you may proceed to highlight your statement as you desire.

[The prepared statement of Millard Ruud follows:]

PREPARED STATEMENT OF MILLARD H. RUUD, EXECUTIVE DIRECTOR, ASSOCIATION OF AMERICAN LAW SCHOOLS

I am Millard H. Ruud, Executive Director of the Association of American Law Schools and for 30 years a member of the law faculty of the University of Texas. While a member of that faculty I served for about 15 years as the chairperson of the school's admissions committee and on various committees of the Law School Admission Council, the national organization that owns the Law School Admission Test and manages the law school admission program. I served as its chairman from 1966 to 1969. Today I appear as the representative of the Association of American Law Schools (AALS).

The Association of American Law Schools was founded in 1900 and is composed of 136 member law schools, all of which are approved by the American Bar Association. The Association is recognized as a national accrediting body by the Council on Postsecondary Accreditation, a non-profit, non-governmental association. As stated in its bylaws, the purpose of our Association is the "improvement of the legal profession through legal education."

The Law School Admission Council (LSAC) is a separate entity. It was established in 1948 and was incorporated as a non-profit educational corporation under the Education Laws of New York in 1968. It is governed by a Board of Trustees; its president (Dean L. Orin Slagle of the College of Law, Ohio State University) is elected by representatives of the member law schools. Presently, the Law School Admission Test (LSAT), Law School Data Assembly Service (LSDAS) and other admissions research and services are administered largely by Educational Testing Service (ETS) under a contract between it and the LSAC. Recently it established under Delaware laws a subsidiary, Law School Admission Services, Inc. (LSAS), a non-profit corporation; LSAS will administer the LSDAS and other services in the near future. LSAC and not ETS has the policy responsibility for the law school admission services.

The operating and institutional relationships of LSAC and AALS are excellent. While we are not unmindful of unsympathetic of the substantial problems that certain of these legislative proposals have for LSAC, the testimony that I offer today is offered from the perspective of our member law schools. I represent their concerns for their applicants and potential applicants and their programs of legal education.

A distinguished legal educator has observed that the decision as to whom should be taught is as important to a school's program of legal education as the decision as to who should teach. As I will make more clear later, our member schools have found LSAT scores to be a significant and important item of information about applicants that assists them in making sound admissions decisions.

In my testimony I will address the important issues presented by the two bills before the committee in terms of applicants for admission to law school, the LSAT, LSAC services, admissions procedures and criteria of law schools, programs of legal education, the legal profession, and the public interest in legal education and the legal profession. I do this because I represent AALS and its 136 schools and because this is the part of the postsecondary education admission world that I know well.

I appreciate that the tight schedule of this committee permitted it to invite to present testimony only one association that represents educational institutions and that does not have a direct interest in any admissions tests. I know that you have invited the other graduate and professional school associations and institutional associations to submit comment on the two bills that are the subject of this hearing. I recognize, however, that I have in a sense a special responsibility to bring to you the perceptions of the host of admissions committees throughout postsecondary education that are some of the very important consumers of admission test scores and admission test services. I have been in communication with a number of these associations and I understand to some extent their concerns. As is appropriate, I will try to reflect those in this testimony and will try in answer to the questions you may ask to reflect their perspectives to the extent that I can.

My review of H.R. 3564 by Congressman Gibbons and H.R. 4949 by Congressman Weiss, Congresswoman Chisholm and Congressman Miller has disclosed three important purposes or themes, all of which are consistent with the interests of our association and are purposes that responsible educators and testing agencies have sought to serve. Let me turn my attention now to these policy objectives of the bills before you.

Wide dissemination of information about the content and form of admissions tests and of test taking strategies.—A very important purpose of the two bills before you and of the legislation enacted in California and New York is that of providing potential test takers with as much information about the admission test as possible so as to place all who take the test on an equal footing. The special role that the admission test score plays in admissions decisions is that of providing an impartial and nationally uniform basis for judging applicants. None of the other information that law school admissions committees have about their applicants has a common and impartial basis. The grade point averages were earned at several hundred different colleges and universities, the letters of recommendation are written by several thousand different persons, the interviews are conducted by a number of different individuals, and the work experiences are at a great veriety of tasks. In short, every other item of information about the applicants is on a perhaps somewhat comparable but not common base. If the admission test score is truly to have an impartial and common basis, it is essential that those who take the test be placed on as nearly a common basis as is possible.

Much of Section 6 of H.R. 3564 and some of Section 3 of H.R. 4949 are designed to serve this purpose.

It may be useful to you if I describe briefly the steps being taken by LSAC to serve this interest. This will permit you to determine what governmental regulation, if any, is needed and, if so, what the shape of it should be. About 500,000 copies of the Law School Admission Bulletin are distributed without cost each year to persons interested in taking the LSAT directly and through prelaw advisors on college campuses, law school admissions offices, career counseling offices and others. See Attachment A for a copy.[1] In addition to describing how to register for the LSAT and LSDAS, where the test is administered and the like, this bulletin describes the purposes and content of the test. A retired form of the test with the correct answers for each question is set out.

For each type of question it is explained why a given answer is the correct one and the others are not. General suggestions for test taking are provided. For example, it is made clear that this admission test is scored on the basis of correct answers only and that, therefore, it is best to try to answer every question. Suggestions for preparation for the test are given. Detailed information about scores obtained by persons taking the test a second time is provided, along with information as to how law schools treat multiple scores.

The concern for the potential test taker does not stop with the test. AALS and LSAC publish an annually revised Prelaw Handbook. This contains information about the legal profession, preparation for law school, the application process and a description of what it is like to be a law student, All law schools provide a two-page description of the salient elements of their educational program and the great bulk of them provide an admissions profile of that year's entering class. This permits a potential applicant to determine her or his probability for admission. This is all designed with the interest of the LSAT test taker and law school applicant in mind. We want the applicant to have as much information as is feasible of the whole process so that she or he may do and choose well.

High quality of admissions tests.—Section 5 of H.R. 4949, Sections 3 and 6 of H.R. 3564 and statements of supporters of this and related legislation express concern about the quality of standardized admissions tests and interest in improving further the quality of standardized postsecondary admissions tests. Our Association shares that interest. We know that LSAC has in its three decades of existence commissioned considerable research to improve the test and in searching for additional forms of testing the aptitude for law study of applicants for admission. The research reports are published in the annual report of the Law School Admisssion Council. These may be found in every law school library and in most of the law school admissions offices. Recently three volumes have been published by LSAC setting out in full the final research reports. Should the committee want copies, LSAC has assured me that they would be happy to supply them.

---

[1] Retained in subcommittee files.

It has been charged by some that admissions testing is "shrouded in secrecy." With respect to the LSAT, the test is "shrouded in secrecy" only from those who are unaware of the research reports or have failed to read and understand them.

LSAC has not relied alone on the research that it has commissioned. It has invited members of the faculty of their member schools to undertake admissions research and has offered to provide financial support for that research.

Another step that has been taken in the effort to maintain and improve the quality of the LSAT has been a publication policy of ETS, warmly supported by LSAC. The professional staff of ETS are encouraged to convert the research that they undertake for client agencies, such as LSAC, into articles publishable in their learned journals. In this way, the ETS professionals are subject to the criticism of their peers—the clinical psychologists, psychometricians and other professionals based in the public and private sector.

Public accountability of the admission process.—H.R. 3564 and H.R. 4949 both recognize the substantial public interest in admissions tests and their use. Section 7 of H.R. 4949 seems to address particularly the issue of accountability.

I can assure you that the LSAC has throughout its history been very conscious of its stewardship responsibility to ensure that the interests of the major consumers of its service, our member schools and their applicants are taken fully into account. To assist it in discharging this responsibility, the LSAC has invited our officers and professional staff to attend meetings of its Board of Trustees and its standing committees. Our responsibility at those meetings has been to provide an additional voice for these interests. By this I do not mean in any way to suggest that the LSAC leadership has not also been very attentive to these interests.

The LSAC publishes an annual report that sets out the minutes of the meetings of its Board of Trustees, and of all of its committees, the research reports, and extensive financial information. These are distributed to all of the law schools and are available to the public.

LSAC takes a number of steps to explain to admissions committees and other admissions personnel the test, its uses and recent research results. At its annual June meeting one or more admissions persons from the member schools participate in one and one-half days of workshops. A week-long workshop is given annually to 60 to 75 admissions persons. Five one day regional admission workshops are given annually. By periodic newsletters to admissions persons current developments are described. Each admissions office is provided with copies of the Legal Affairs Manual and Operations Reference Book. LSAC has just distributed its Cautionary Policies, which is an excellent guide to use of the LSAT scores and other admission criteria. These and other means are taken to assist the law schools to make appropriate use of the LSAT scores and other admission criteria and to report to them on LSAC's work on their behalf. Our Section on Prelegal Education and Admissions to Law School through its annual program and otherwise assists in these efforts.

The LSAC has recently started meeting annually with the deans of their member schools for the purpose of reporting to the deans on LSAC's programs asking them for suggestions of ways in which the program of testing and admissions services may be further improved.

For a number of years the LSAC has held periodic conferences with prelaw advisors for the purpose of hearing from them ways in which the program may better serve their interests and those of their students. In addition, the officers, committee chairpersons, and professional staff have attended regional meetings of the prelaw advisors for the purpose or reporting on their work and hearing their comments.

With respect to the test takers, the LSAC has over the years conducted periodic surveys for the purpose of receiving their comments on how the test and test services may be improved. One was just completed.

Abolition of standardized admissions tests.—As I have read over the past year the statements in support of legislation like that before the Committee, I have come to believe the purpose of some of the supporters of test disclosure legislation is the elimination of the use of standardized admissions tests. The apparent strategy is to require publication of each form of the test shortly after its use. It is assumed that the testing agencies will be unable to continue to produce sufficient number of new forms of the quality equal at least to the present tests. This would then bring these programs to an end. Failing to secure abolition by these means, the apparent hope of some is that the quality of the tests may be so reduced as the product of a greatly increased production required by disclosure that the test scores will have significantly less value and so will make less contribution to the prediction of probable academic performance of applicants.

Our Association emphatically opposes this objective. Justice Oliver Wendell Holmes observed that often a page of history is worth more than a volume of logic. The history of the Law School Admission Test is this. In 1947 Professor Willis L. M. Reese of Columbia University School of Law and representatives of another half dozen law schools went to the newly established ETS to ask President Henry Chancey whether it would be willing to try to establish a test that these law schools might use to assist the schools in making better predictions of academic success. These legal educators wished to treat more equally applicants with college records from a variety of schools and to reduce significantly the first year attrition rates by doing a better job of selecting those with more substantial probabilities for success. The LSAT was not something that Educational Testing Service devised and then marketed to the law schools. The idea was ours and we enlisted the assistance of the testing agency.

Our member schools and the other nationally approved law schools have three decades of experience with the LSAT. That experience had demonstrated the validity and reliability of the test and the substantial contribution it can make to law school admissions. Our Association has been so persuaded of the importance of an admissions test score to sound admissions that the bylaws of the Association provide: "A member school should determine whether an applicant is adequately equipped for the study of law on the basis of an undergraduate record, an admission test score, training, experience and other indicia of aptitude for the study of law." Standard 503 of the Standards for Approval of Law Schools by the American Bar Association similarly urges the use of an admissions test and requires a school that is not using the LSAT to "establish that it is using an acceptable test." The process of our Association adopting bylaws and the process of the American Bar Association adopting its standards is one that ensures deliberation and full input of interested and effected parties. In short, admissions tests are included in the accreditation criteria for sound reasons.

It is commonly said that admissions tests are culturally biased. In a strictly technical sense, this is true and it is necessarily true. Linguists will tell you that all languages reflect the culture of those who use them. The question then is not whether the LSAT is, strictly speaking, culturally biased, but whether it is fair to applicants from racial and ethnic minority backgrounds. The purpose of the test is to predict academic performance. If the LSAT underpredicts the potential for black applicants generally, for example, then the test should not be used in judging the law school potential of these applicants.

Beginning in 1963 LSAC has conducted studies of the LSAT to determine whether the prediction formula developed upon the basis of the validity studies works as well for black and other minority applicants as it does for white applicants or applicants generally. The findings of these studies, from the first one published in 1966 to the most recent one published in 1978, demonstrate that black and Mexican American applicants predict in the same way as other applicants. In short, a school does not discriminate against these minority applicants when it uses its conventional prediction formula to establish a minority applicant's potential for law study.

This has not precluded law schools from practicing affirmative action in admissions. A study of the fall 1976 entering class suggests that had the law schools not taken into account race or ethnic background in making admissions decisions, the number of blacks receiving offers of admissions would have been about 60 percent less than they were and Mexican Americans or Chicanos about 40 percent less than they were. Applications and Admission to ABA Accredited Law Schools: An Analysis of National Data for the Class Entering in the Fall 1976 ("Law School Admission Research," May 1977).

In summary, our member schools find the LSAT score an important and reliable datum concerning the potential of their applicants for law study. We would object very much to any efforts to deprive them of this.

Let me turn now to specific provisions of H.R. 3564 and H.R. 4949.

Test disclosure requirement.—The requirement in the bills before you that presents the most difficulties from the perspective of our member schools in Section 5 of H.R. 4949, which requires the public filing, among other things, of each test and the acceptable answers for each question and requires the test agency to send to each person who took the test the same information plus the "test subject's individual answer sheet." Fortunately for the future of admissions tests only those questions used to calculate the test subjects raw score need be disclosed. As you may know, each form of the LSAT contains some questions whose answers do not go into calculating the candidate's score; these are questions being protested. Test questions are developed by educated, trained and experienced item writers according to established guidelines and are subject to careful review. However, they are not used until they have been proven effective and reliable through this pretesting.

The disclosure requirement on its face seems eminently fair and reasonable. Let me provide you with some background to permit you to make an informed judgment about this. The present practice of LSAC is to retire any form of the test that is compromised or very probably compromised. Thus, if there is an established unauthorized removal of a test and if there is a possibility that some distribution has been made of the test, the form is retired. This is done so that each person who takes the LSAT is on as equal a footing as possible. If there is a probability that some test takers have seen the form of the test, those test takers have an advantage. Therefore, the policy is to retire a compromised form of the LSAT. As a matter of practice, a test form is retired after its use in three regular Saturday administrations, three Monday administrations and some special administrations. The cost of producing a form of the LSAT may thus be amortized over six scheduled administrations plus a moderate number of special administrations. The proposed disclosure requirement would require the retirement of the test after one administration. It is not clear just what should be done about special administrations, but more about that later. This means that the cost of the production of a test form would need to be charged entirely against a single administration. Obviously, this substantially increases the cost for that administration.

It is my understanding that the cost of producing a single form of the LSAT is now between $80,000 and $120,000. Presently the cost of retiring about three forms a year are charged against the approximately 128,000 administrations of the test.

Increased cost is only one of the difficulties presented by disclosure requirement. The most troubling is whether there is an indefinite number of questions that may be written so that the testing agency has the capacity to continue to produce more test questions. For those admissions tests that are anchored in a base of information there is a limited capacity to produce additional questions. With respect to that form of admissions tests, the disclosure requirement is a death sentence. I understand this to be the case with respect to the New Medical College Aptitude Test and with respect to the Spatial Perception Test of the American Dental Association. It may not be the case with respect to the LSAT. While it is anchored in information of a kind, the information is of a general sort and the principal purpose of the test is to test for certain skills and capacities. There may be practical limits, but the limits are not so severe.

In the short run there is the limitation of the number of educated, trained and experienced personnel to write the test items. In the short run, too, there is the time needed to draft and review test items, subject them to pretesting, review the results of the pretesting to determine the reliability and validity of the test items to determine which may be used and to determine their different levels of difficulty. In the long run, if may be possible to produce a sufficient quantity of tests questions of current quality to permit disclosure and retirement after a single use of the test items. That this is so is still not clear.

If it were to be done, a number of adjustments would need to be made. To hold the increased cost to the testing program and thus to the test taker within some reasonable limits, it will be necessary to reduce the number of administrations of the test. In examining costs it must be remembered that the only source of income to support the LSAC's program is the fees paid by persons taking the LSAT and registering for the LSDAS. Any additional costs are costs that the Council can pass on to no one other than those considering making application to law school.

A difficulty that compliance with the disclosure requirement will present our member schools and their potential applicants flows from the reduced frequency of administration of the LSAT. To hold down the annual cost of test construction and to insure production of quality tests, the testing agencies will need to reduce the number of tests that must be retired. This can be done by using a single form instead of one to three forms in a given administration and by reducing the number of administrations each year.

The prime candidates for elimination are the low volume test administrations. The test is administered on five Mondays a year to accommodate those whose religion recognizes Saturday as their Sabbath. Should the Monday administrations be eliminated or at least reduced substantially? Why is this necessary, you may ask. After all, H.R. 4949 does not require publication of the test given on Saturday until 30 days after the scores have been reported. Therefore, the Saturday test is not made public until six to eight weeks after the Monday administration. The concern about using the same form on Monday as on Saturday is that a few persons might take the test on Saturday under an assumed name and on Monday under their own name. In this way they would have an opportunity to practice taking the test on the very test form that they are going to take "for real."

One response could be to use to a new form but offer the test on Monday only once a year. Another response would be to offer as many Monday administrations as Saturday administrations and ask the schools to assume that the special religious commitment of these test takers will give them assurance that there is no or virtually no test misconduct. That some schools would value the Monday LSAT scores less in these cases is, however, a possibility. In either response, the service to law school applicants and our member schools would be reduced as a consequence.

I suggest that the Committee consider exempting the Monday and other special administrations from the requirement to disclose the test questions.

Presently the LSAT may be administered in this country and around the world at virtually any date in addition to the ten scheduled Saturday and Monday administrations per year. Armed forces personnel stationed around the globe may arrange to have the test administered to them wherever they are and at any time. The Peace Corps worker who returned from Africa just after the February administration and needs to have a test score earlier than that which would be produced by an April test to have an opportunity to enter law school in the fall can arrange, by paying a special fee, to have a special administration. Some handicapped persons want special administrations. It is not readily apparent to me or to the leadership of the LSAC Council how this valuable program of special administrations can be continued. The suggestion that the form of the test used at the most recent general administration and that has not yet been published be used in these special administrations encounters virtually the same problems as those suggested for the Monday administrations of the test form used on Saturday. Unless H.R. 4949 is amended to exclude the special administrations from the disclosure requirements as now stated, it appears that there is no practical way to continue the program of special administration. That would be a shame.

What are the implications for applicants to law school and law schools in the understandable response to the testing agencies and test sponsors to the disclosure—increased cost and decreased frequency of test administration? Increased cost would, of course, have its greatest impact on the economically disadvantaged. The minority applicants are disproportionately represented in the economically disadvantaged. It would be a national tragedy if these increased costs caused a significant number of economically disadvantaged and minority persons to fail to take the LSAT and thus fail to complete their applications for admissions to law school. The LSAC has a modest program of waiving fees for the severely economically disadvantaged. The cost of that program is, of course, borne by those who pay the fees. The increased costs caused by the disclosure requirements will significantly increase the cost of the test fee waiver program at its present level and so would make it difficult to increase significantly the level of that program.

The decreased frequency of administration of the test caused by the disclosure requirement causes us even greater concern. We understand that some of the support for the bill before you and similar bills introduced in state legislatures came from legislators and citizens with a special interest in the minority applicant for law and other professional schools. Our Association shares that interest. Our Association was a founding member of the Council on Legal Education Opportunity, which has for the past decade played an important role in increasing the enrollment of minority and disadvantaged persons in the nation's law schools. It is a federally funded program. We filed an amicus curae brief in the *Bakke* case in support of the general position of the University of California, both with the Supreme Court of California and the Supreme Court of the United States.

Our judgment is that the disclosure requirement in its present form would in fact harm and not help the minority applicants for law school. I have mentioned the probable impact of the increased costs of the test. The less frequent availability of the test is likely to affect those persons who are newest to access in higher education and whose undergraduate colleges provide less structure and less reinforcement for advanced planning for the complex law school admission process. The lead time for making application for law schools is the longest that it has been. To have the full opportunity to be considered, an applicant should have completed his or her application, including the taking of the LSAT, by March or April 1. I am told by deans of admissions that minority applicants as a group tend to make their decisions to take the test and apply later in the admissions cycle than applicants generally. The reduction in the number of test administrations would seem to disproportionately disadvantage the minority applicant.

Release of test subject's individual answer sheet.—H.R. 4949 Section 5(a)(2)(B) requires the testing agency, at the test subject's request, to send to the test subject the "test subject's individual answer sheet," among other things. I would strongly urge the committee to amend that to read "a copy of the test subject's individual answer sheet." This would make it clear that the test agency should keep the original. If it must surrender the original, then we would be concerned about the resolution of disputes about scoring. If the test agency had a copy and that copy showed that the test subject had answered certain questions differently from that indicated on the original, which the test subject offers in evidence to the court, the best evidence rule might require the court to admit the original in its present form as evidence and to refuse to admit the copy retained by the test agency. A moment's reflection suggests to you and to me a potentiality for successful alteration that this rule offers. I cannot believe that the drafters of the bill intended this result. It is very probably an oversight.

There are a number of technical problems presented for the testing program and our schools by the specific provisions of the bills before you. It may well be that these are largely the product of the drafters not understanding fully the practical consequences of the bill's formulation of the rule. Let me provide you with some examples of the kinds of provisions to which I have reference. It would trespass too much on the committee's time to go through them section by section. I am certainly willing to make myself available to the committee's staff to go over these matters in detail. I suggest to the committee that it may wish to direct its staff to confer also with representatives of agencies like the Law School Admission Council and testing agencies like the American College Testing Program and Educational Testing Service for the same purpose.

Score generally required for admission.—Section 6(b)(3) of H.R. 3564 requires the testing agency at the request of test subject, to notify her or him of the "score which is generally required for admission to institutions of higher education." With respect to the LSAT in law school admissions, there is literally no such score. First of all, the competition for seats in our 136 member schools varies. Secondly, the LSAT score and the cumulative grade point average are the two quantitative factors usually used to predict law school performance or to develop an admission index. The LSAT score needed to receive favorable consideration at a given school obviously varies with the applicant's grade point average.

The interest this provision apparently addresses is now dealt with in the law school admission program by the publication in the Prelaw Handbook of a profile of the most recently admitted entering class. This may more directly address the interest of the test taker. She or he tends not to be interested in institutions of higher education generally but in specific institutions. Is it not more appropriate that the applicant obtain that information directly from the institutions in question? It may be inappropriate to place on the testing agencies and test sponsors the responsibility of getting that information from the institutions. Our Association and the LSAC respect the autonomy of the member law schools and want to exercise over them only that authority that is absolutely necessary, not simply potentially desirable.

Subsequent test scores.—Section 6 of H.R. 4949 provides, among other things, that the test agency with permission of the test subject release only "all previous scores" to designated law school. Research has informed our member schools that the best way to treat an applicant who has taken the LSAT several times is to average all scores, unless there are special circumstances, such as illness, that dictate that a test score should be ignored. Section 6 would preclude the testing agency from sending our member schools the second and subsequent scores unless expressly authorized by the test subject. Why should the government deprive our schools of this valuable information? Again, this may be simply a drafting error. The testing agencies certainly should be authorized to submit to the schools that the test subject has asked scores to be sent to all scores a person earns.

Disclosures to potential test registrants.—Some of the information that Section 3 requires a test agency to disclose is different for each of our member schools. The apparent assumption of the drafters is that it is the same. For example, Subsection 3 asks that "the correlation between test scores and future success in schools" be provided. The correlation varies. Does H.R. 4949 contemplate publishing this for all our 136 member schools plus the 32 additional ABA approved schools? Also, is any recognition to be given to institutional privacy?

Section 3 also requires the test agency to explain "the correlation between test scores and success in the career for which admission is sought." The LSAT is not designed to predict success in the legal profession, but rather the probability of success as a law student. The LSAC literature contains cautionary statements to law firms and other employers of law graduates that LSAT scores should not be used for this purpose. Law school grades are a much more appropriate criteria.

There is a larger problem. What for these purposes is "success?" Is it measured by income? By merely gaining admission to the bar? I am confident that the drafters had little understanding of what they were asking for by this innocent looking ten word clause. The cost of the research to provide a preliminary answer to this question is conservatively estimated at between $1 million and $2 million. Should the LSAT test takers be required to pay for this? Will they find the results worth the cost?

Subsection 3 would require the Law School Admission Bulletin to contain "a comparison of the average score and percentiles of test subjects by major income groups." The LSAC program does not now gather information concerning the income of the test subject of her or his family. Does this provision require the LSAC to require each test taker to provide this information so that the studies called for may be made and this information reported? Are there not questions of privacy? again, is the cost of this effort worth the candle?

Regulation of the use of standardized admission test scores.—Some of the public statements made in support of the testing regulation legislation like that before this committee has stated as one of the principal purposes of the legislation to be to cure misuses of standardized admission test scores. H.R. 4949 as introduced contained a version of Section 8 that required an explanation of the use of test scores as criterion for selection of students by the undergraduate, graduate and professsional programs of colleges and universities. If I am correctly informed, this provision has now been withdrawn. If it has not, then it would seem to regulate through the means of publicity the use of test scores. The concern that I have with the provision is that this publication requirement is likely to introduce greater rigidity in admissions. Litigation over admissions would test admissions actions against the institution's stated formula. Aware of this, the admissions committee is more likely to follow literally the stated formula. While this gives an appearane of greater fairness, it also reduces the role of informed judgment in making admissions decisions.

If I am correctly informed, then the only way in which either of these bills regulates the use of the scores by our law schools is through the hope that some may have for the disclosure provisions of H.R. 4949. As I indicated earlier, it is quite clear that the hope of some supporters of the required publication of each test after it has been given is that standardized tests will be driven out of business. We are, as I said earlier, emphatically opposed to that.

Financial disclosure.—LSAC in its annual report, which is distributed to our member schools and other law schools and which is available in law school libraries all around the country, provides all of the information requested. The committee may wish to ask its staff to consult with representatives of LSAC and other test agencies to determine whether specifying that this report and test volume report must be made within "120 days after the close of the testing year" presents problems and whether the fiscal year instead of the testing year should be specified in setting the reporting date. Clearly the revenue and volume figures should be in terms of the testing year.

The prices for all of the testing and admissions services by LSAC, are set at levels to generate sufficient income to pay the costs of the program. By estimating the probable volume for the test year and the costs of providing the service under its contract with the Educational Testing Service, the Board of Trustees of LSAC sets the various fees for its services. To demonstrate that this is not an exact science, I should report to you that expenditures exceeded income in fiscal 1979 by over $600,000. Unexpected decreases in volume caused the deficit.

In conclusion, let me express may appreciation to the Committee for giving me this opportunity to present the views of our Association. As I have indicated, we share most of the policy objectives of these two bills. Our concerns are with the means chosen to serve the ends.

Thank you.

Mr. RUDD. One other step that we take for this purpose is the publication, and in this case, "we" is correct.

Our association and the Law School Admission Council jointly publish the Prelaw Handbook that has information on all 168 ABA approved schools with admissions profiles, so that persons considering application to a given school can get some notion of what the probabilities are of gaining admission.

A second purpose that I see in section 5 of H.R. 4949 and sections 3 and 6 of H.R. 3564 is a concern about the quality of standardized admission tests, and an interest in improving further the quality of the tests.

Our association certainly shares that interest. We know of three decades of commission research by the Law School Admission Council that has sought that very end.

Here are three volumes summarizing that research. It has been charged by some that admissions testing is shrouded in secrecy. With respect to the LSAT at least the test is shrouded in secrecy only from those who are apparently unaware of this information or have failed to read and understand it.

A third purpose sought is public accountability of the admissions process. We certainly share that.

The Law School Admission Council, I know, has been very conscious throughout its history of the stewardship responsibility it has in managing the resources which, as someone indicated earlier, comes entirely from the students who pay the fees and the leadership has been conscious of that.

Among the ways that they try to serve the public accountability purpose is the publication of an annual report, and those are available in any law school library in the country, and they have complete minutes of the governing board of all its standing and special committees are there.

There is nothing secret there. It may be found and used.

Like what you heard about the tests, some of the tests, from witnesses, the Law School Admission Council takes a number of steps to try to explain to admissions committees and other admissions personnel the test, its limitations, its strengths, how best to use it by conferences of a variety of a kind, its annual meeting, its annual workshop for a smaller group, regional meetings.

Our association has a section on prelegal education and admissions to law school, and through its annual program and otherwise seeks to serve that purpose over the past year.

I have read the statements by some who suppose the legislation like that before you, and I have come to believe that some, I am confident none of the authors of the legislation have this purpose, but I have come to believe that the purpose of some of the supporters of the test disclosure legislation is elimination of the use of standardized admissions tests.

As I understand the strategy, or as it appears to me to be, is to require publication of each form of the test shortly after it is used, and assuming that the testing agencies will be unable to continue to produce enough new forms and tests of equal quality to at least the present ones, this would bring the program to an end.

If that doesn't work, then apparently the hope is that the quality of what can be produced may be so reduced as the product of the greatly increased production required by disclosure of the test scores will have a significantly less value, and so will make less contribution to the prediction of probable academic performance.

To the extent that that is anybody's purpose, our association emphatically opposes that object.

Mr. Weiss. Would you hold your thought right there?

The second bells have sounded.

We now have about 9 minutes to get our vote in.

We will return in about 10 or 12 minutes.

Thank you.

Chairman PERKINS. The gentleman from the law schools, Mr. Ruud; you may continue.

Mr. RUUD. When the committee recessed I was making the point to which some support for the legislation before the committee has been aimed, at the abolition of standardized tests. Our association opposed that objective. Justice Holmes once said: "A page of history is worth a volume of logic." In this case, it is good to recall that the law school admission test was not something devised by a few. A group of people from our member schools went to the fledgling Educational Testing Service in 1947 and said, can you help us, can you devise a test that will be helpful to us in making better admissions decisions than we are now making and opening up the possibility of considering applications from undergraduate institutions with which we have had no experience?

Our association considers the LSAT or some standardized admission test vital and very valuable to the admissions process, as manifested by one of our rules of membership or accreditation criteria, which provides:

> A Member school should determine whether an applicant is adequately equipped for the study of law on the basis of an undergraduate record, an admission test score, training, experience, and other indicia of aptitude for the study of law.

It is commonly said that admissions tests are culturally biased. In a strictly technical sense, this is true and it is necessarily true. Linguists will tell you that all languages reflect the culture of those who use them. The question then is not whether the LSAT is, strictly speaking, culturally biased, but whether it is fair to applicants from racial and ethnic minority backgrounds. The purpose of the test is to predict academic performance. If the LSAT underpredicts the potential for black applicants generally, for example, then the test should not be used in judging the law school potential of these applicants.

All the studies—and these are reported in these three volumes that are before me, the volumes setting forth the research of the Law School Admission Council—the findings of these studies, from the first one published in 1966 to the most recent one published in 1978, demonstrate that black and Mexican American applicants predict in the same way as other applicants. In short, a school does not discriminate against these minority applicants when it uses its conventional prediction formula to estimate a minority applicant's potential for law study.

In summary, our member schools find the LSAT score an important and reliable data concerning the potential of their applicants for law study. We would object very much to any legislation that had the effect of depriving our schools of this or reducing substantially the access of applicants to LSAT services.

Let me turn now to some of the specific provisions of the two bills. The test disclosure requirement is one that presents the most difficulties from the perspective of our member schools. The disclosure requirement on its face seems eminently fair and reasonable. Let me provide you with some background to permit you and the members of the committee to make a judgment about this. The present practice, as I understand it, from LSAT's, they retire any form of the test that is compromised or very probably compromised. If there is established or unauthorized taking or removal of a test or if there is a possibility distribution has been made of the test, the form is retired. That is done so that those who might have a copy of the test do not have a special advantage.

The test form as a matter of practice is now retired after three regular Saturday administrations, three Monday administrations, and some special administrations. The cost of producing that form is then amortized over those six nationally scheduled administrations, plus the moderate number of special administrations. The proposed disclosure requirement in H.R. 4949 would require the retirement of the test after a single administration.

It is not clear just what should be done about special administrations and how the testing agencies can accommodate to this.

It is my understanding the cost of producing a single form presently of the LSAT is between $80,000 and $120,000. Presently the cost of retiring about three forms a year are charged against approximately 128,000 administrations of the test.

Increased cost is only one of the difficulties presented by the disclosure requirement. The most troubling is the question of whether there is an indefinite number of questions that may be written. For those admissions tests that are anchored in a base of information there is a limited capacity to produce additional questions. With respect to that form of admissions tests, the disclosure requirement is a death sentence. I understand this to be the case with respect to the new medical college aptitude test and with respect to the spatial perception test of the American Dental Association. It may not be the case with respect to the LSAT. While it is anchored in information of a kind, the information is of a general sort and the principal purpose of the test is to test for certain skills and capacities. There may be practical limits, but the limits are not so severe.

The principal purpose of the test is to test for certain skills and abilities.

There are practical limits in the short run, but the limits on production of test items seem not to be as severe as it is in other kinds.

In the short run, there is a limitation on the number of educated, trained personnel one can employ, get on the job to write the test item.

In the short run, there is the problem of the time needed, the lead time needed to draft and review and put through pretest and so on.

In the long run, it may be possible to produce a sufficient quantity of test questions to permit disclosure and retirement after single use of the test.

After I prepared this testimony, I was told if the length of the LSAT was from 3½ to 7 hours, they would be able to do enough pretesting to be able to produce the quantity of tests needed.

If LSAT can accommodate to this disclosure in terms of maintaining the quality of the test, a number of adjustments need to be made, an adjustment that involves reduction in the number of administrations of the test. That is one way to hold down the annual cost. You then have to retire fewer tests in a given year.

The prime candidates for elimination are of course the low-volume test administrations. One of those low-volume groups are the five Monday administrations that are administered to those who honor Saturday as their Sabbath. How should they respond? Eliminate perhaps all but one?

Why is this necessary, you may ask? After all, H.R. 4949 does not require publication of the test given on Saturday until 30 days after the scores have been reported. Therefore, the Saturday test is not made public until 6 to 8 weeks after the Monday administration. The concern about using the same form on Monday as on Saturday is that a few persons might take the test on Saturday under an assumed name and on Monday under their own name. In this way they would have an opportunity to practice taking the test on the very test form that they are going to take for real.

One response could be to use a new form but offer the test on Monday only once a year. Another response would be to offer as many Monday administrations as Saturday administrations and ask the schools to assume that the special religious commitment of these test takers will give them assurance that there is no or virtually no test misconduct. That some schools would value the Monday LSAT scores less in these cases is, however, a possibility. In either response, the service to law school applicants and our member schools would be reduced as a consequence.

I suggest that the committee consider exempting the Monday and other special administrations from the requirement to disclose the test questions.

Presently the LSAT may be administered in this country and around the world at virtually any date in addition to the 10 scheduled Saturday and Monday administrations per year. Armed Forces personnel stationed around the globe may arrange to have the test administered to them wherever they are at any time. The Peace Corps worker who returned from Africa just after the February administration and needs to have a test score earlier than that which would be produced by an April test to have an opportunity to enter law school in the fall can arrange, by paying a special fee, to have a special administration. Some handicapped persons want special administrations. It is not readily apparent to me or to the leadership of the LSAC Council how this valuable program of special administrations can be continued. The suggestion that the form of the test used at the most recent general administration and that has not yet been published be used in these special administrations encounters virtually the same problems as those suggested for the Monday administrations of the test form used on Saturday. Unless H.R. 4949 is amended to exclude the special administrations from the disclosure requirements as now stated, it appears that there is no practical way to continue the program of special administration. That would be a shame.

What are the implications for applicants to law school and law sponsors to the disclosure—increased cost and decreased frequency of test administration? Increased cost would, of course, have its greatest impact on the economically disadvantaged. The minority applicants are disproportionately represented in the economically disadvantaged. It would be a national tragedy if these increased costs caused a significant number of economically disadvantaged and minority persons to fail to take the LSAT and thus fail to complete their applications for admission to law school. The LSAC has a modest program of waiving fees for the severely economically disadvantaged. The cost of that program is, of course, borne by those who pay the fees. The increased costs caused by the disclosure requirements will significantly increase the cost of the test fee waiver program at its present level and so would make it difficult to increase significantly the level of that program.

The decreased frequency of administration of the test caused by the disclosure requirement causes us even greater concern. We understand that some of the support for the bill before you and similar bills introduced in State legislatures came from legislators and citizens with a special interest in the minority applicant for law and other professional schools. Our association shares that interest. Our association was a founding member of the Council on Legal Education Opportunity, which has for the past decade played an important role in increasing the enrollment of minority and disadvantaged persons in the Nation's law schools. It is a federally funded program. We filed an amicus curiae brief in the *Bakke* case in support of the general position of the University of California, both with the Supreme Court of California and the Supreme Court of the United States.

Our judgment is that the disclosure requirement in its present form would in fact harm and not help the minority applicants for law school. I have mentioned the probable impact of the increased costs of the test. The less frequent availability of the test is likely to affect those persons who are newest to access in higher education and whose undergraduate colleges provide less structure and less reinforcement for advanced planning for the complex law school admission process. The lead time for making application for law schools is the longest that it has been. To have the full opportunity to be considered, an applicant should have completed his or her application, including the taking of the LSAT, by March or April 1. I am told by deans of admissions that minority applicants as a group tend to make their decisions to take the test and apply later in the admissions cycle than applicants generally. The reduction in the number of test administrations would seem to disproportionately disadvantage the minority applicant.

There are a number of technical problems relating to the testing program and our school by the specific provisions of the bill before you. It may be these are largely the product of the drafters not understanding the bill's formulation of the rule.

Let me provide the committee with some examples of these kinds of provisions. Perhaps a few examples of this would illustrate the problem.

RELEASE OF TEST SUBJECT'S INDIVIDUAL ANSWER SHEET

H.R. 4949 section 5(a)(2)(B) requires the testing agency, at the test subject's request, to send to the test subject the "test subject's individual answer sheet," among other things. I would strongly urge the committee to amend that to read "a copy of the test subject's individual answer sheet." This would make it clear that the test agency should keep the original. If it must surrender the original, then we would be concerned about the resolution of disputes about scoring. If the test agency had a copy and that copy showed that the test subject had answered certain questions differently from that indicated on the original, which the test subject offers in evidence to the court, the best-evidence rule might require the court to admit the original in its present form as evidence and to refuse to admit the copy retained by the test agency. A moment's reflection suggests to you and to me a potentiality for successful alteration that this rule offers. I cannot believe that the drafters of the bill intended this result. It is very probably an oversight.

Mr. WEISS. You are right.

Mr. RUUD. My responsibility, Congressman, is to tell you where we live and what the bill in its present form is likely to do to us. We share common interests and we want to do the right thing for the folks.

Section 6 of H.R. 4949 provides among other things that the test agency, with permission of the test subject, may release all previous scores to the designated law school.

Research has informed our member schools that the best way to treat an applicant who has taken the LSAT several times is to average all scores unless there may be special circumstances to dictate a given score should be ignored.

Section 6 precludes by law the testing agency from sending to our member schools the second and subsequent scores unless expressly authorized by the test subject. Why should the Government deprive our schools of this valuable information? Again, this may be simply a drafting error. The testing agencies certainly should be authorized to submit to the schools that the test subject has asked scores to be sent to, all scores a person earns.

There are several items in section 3.

Section 3 also requires the test agency to explain "the correlation between test scores and success in the career for which admission is sought." The LSAT is not designed to predict success in the legal profession, but rather the probability of success as a law student. The LSAC literature contains cautionary statements to law firms and other employers of law graduates that LSAT scores should not be used for this purpose. Law school grades are a much more appropriate criteria.

There is a larger problem. What for these purposes is "success"? Is it measured by income? By merely gaining admission to the bar? I am confident that the drafters had little understanding of what they were asking for by this innocent looking 10-word clause. The cost of the research to provide a preliminary answer to this question is conservatively estimated at between $1 million and $2 million. Should the LSAT test takers be required to pay for this? Will they find the results worth the cost?

Subsection 3 would require the Law School Admission Bulletin to contain "a comparison of the average score and percentiles of test subjects by major income groups." The LSAC program does not now gather information concerning the income of the test subject of her or his family. Does this provision require the LSAC to require each test taker to provide this information so that the studies called for may be made and this information reported? Are there not questions of privacy? Again, is the cost of this effort worth the candle?

Regulation of the Use of Standardized Admission Test Scores. Some of the public statements made in support of the testing regulation legislation like that before this committee have stated as one of the principal purposes of the legislation to be to cure misuses of standardized admission test scores. H.R. 4949 as introduced contained a version of section 8 that required an explanation of the use of test scores as criterion for selection of students by the undergraduate, graduate, and professional programs of colleges and universities. If I am correctly informed, this provision has now been withdrawn. If it has not, then it would seem to regulate through the means of publicity the use of test scores. The concern that I have with the provision is that this publication requirement is likely to introduce greater rigidity in admissions. Litigation over admissions would test admissions actions against the institution's stated formula. Aware of this, the admissions committee is more likely to follow literally the stated formula. While this gives an appearance of greater fairness, it also reduces the role of informed judgment in making admissions decisions.

If I am correctly informed, then the only way in which either of these bills regulates the use of the scores by our law schools is through the hope that some may have for the disclosure provisions of H.R. 4949. As I indicated earlier, it is quite clear that the hope of some supporters of the required publication of each test after it has been given is that standardized tests will be driven out of business. We are, as I said earlier, emphatically opposed to that.

FINANCIAL DISCLOSURE

LSAC in its annual report, which is distributed to our member schools and other law schools and which is available in law school libraries all around the country, provides all of the information requested. The committee may wish to ask its staff to consult with representatives of LSAC and other test agencies to determine whether specifying that this report and test volume report must be made within "120 days after the close of the testing year" presents problems and whether the fiscal year instead of the testing year should be specified in setting the reporting date. Clearly the revenue and volume figures should be in terms of the testing year.

The prices for all of the testing and admissions services by LSAC are set at levels to generate sufficient income to pay the costs of the program. By estimating the probable volume for the test year and the costs of providing the service under its contract with the Educational Testing Service, the board of trustees of LSAC sets the various fees for its services. To demonstrate that this is not an exact science, I should report to you that expenditures exceeded income in fiscal 1979 by over $600,000. Unexpected decreases in volume caused the deficit.

In conclusion, let me express my appreciation to the committee for giving me this opportunity to present the views of our association. As I have indicated, we share most of the policy objectives of these two bills. Our concerns are with the means chosen to serve the ends.

Chairman **Perkins**. We appreciate your testimony this morning, Mr. Ruud.

Our next witness is Mr. Jon Haber, consultant, U.S. Student Association, who is accompanied by Mr. Andre Burnett. Then we will hear from Dr. Roger T. Lennon, Harcourt, Brace, Jovanovich, Inc.

You may proceed.

[The prepared statement of Jon Haber follows:]

Prepared Statement of Jon Haber, U.S. Student Association, Accompanied by Andre Burnett, Director, National Third World Student Coalition, U.S. Student Association

Mr. Chairman and members of the subcommittee, I thank you for the opportunity to appear before you today. My name is Jon Haber and I am representing the United States Student Association (USSA). I was formerly the CoDirector of the University of California Student Lobby and was the principal drafter of the California Truth-in-Testing legislation. I am accompanied by Andre Burnett, Director of USSA's National Third World Student Coalition. We are speaking on behalf of USSA's 3,000,000 student members attending over 350 colleges and universities.

We join many other organizations in strongly endorsing H.R. 4949—the Educational Testing Act of 1979, including the National Parent Teachers Association (PTA), National Education Association (NEA), NAACP, National Public Interest Research Group (PIRG), Coalition of Independent Colleges and University Students (COPUS), the Americans for Democratic Action (ADA), as well as many statewide student associations.

The Educational Testing Act of 1979 is designed to provide more information about standardized tests and promote a wider understanding of these tests amongst students, educators, parents, admissions officials, and the public in general.

Equally important is what this Act does not do. It does not regulate test contents nor testing procedures of the testing companies, nor any aspect of their admissions testing program. Neither does it specify how colleges and universities shall use standardized tests in their admissions programs. This Act is simple disclosure legislation—no more and no less.

It is also important to stress that we are not anti-testing. We do not support the abolition of the testing companies or the tests they make. We simply believe that tests measure a narrow range of characteristics and are therefore subject to misuse and misunderstanding. Through the information provided them by H.R. 4949, students and the public will gain a better understanding of the nature, purposes, uses, and limitations of standardized admissions tests. We also believe that this Act will ensure that tests are used properly—a goal ETS and other testing agencies supported before this Subcommittee yesterday.

Lastly, we have little to say about H.R. 3564 (Gibbons). While we applaud the intent of the legislation, if passed, it would have no impact on the tests, and would not improve or change the amount of information presently provided to students. It essentially codifies the testing companies' current practices.

American colleges and universities rely heavily on standardized tests, often using them as the critical factor in deciding what students to admit to their undergraduate, graduate, and professional programs. Although not covered by this legislation, standardized tests also play an extensive role in determining which students will qualify for financial aid, which students will be awarded advanced placement credit, and ultimately, which students can be professionally licensed after their college training.

I have been involved with the question of standardized testing for three years. I have spoken and met with thousands of students about tests and at this point will share some of their concerns with you.

Students feel that they are victims, not consumers, of standardized tests. They have little choice but to take the test—their college or university has made it a mandatory component of the admissions process. And although the test will be made by one of two testing companies, the student will usually not be allowed to decide whether to take ETS' or ACT's test—the college has already made that choice for them.

Preparing for and taking the test is extremely stressful. Not only are the tests taken during crucial periods in students' lives—when they are deciding what paths to follow after high school, undergraduate or graduate schooling—but also because each knows that their test score may indeed decide their professional and occupational future.

While students do not know the test's margin of error, predictive value, or bias, students are well aware of one fact—their score may mean the difference between being accepted or rejected at most colleges. As a result, many students turn to expensive coaching schools or coaching books in hopes of better test scores.

Students are frustrated at being unable to see their questions and answers. They can never know which questions they answered correctly, or in which subjects they need more work. Perhaps they misnumbered their answer sheet or the computer has made another scoring error.

Students are often resentful when they receive their scores. They often feel that the test did not really reflect their ability.

Many feel that essay tests would have been fairer indicators of their talents.

Students believe that the tests are overpriced. A student applying to seven law schools must spend as much as $35 simply to take the test and have his/her score and grades mailed out. Additionally, this student might pay as much as $140 in school application fees, not to mention the costs of coaching books or coaching schools to prepare for the test.

I am not speaking of a handful of students. Rather, many students feel this. In 1977-78, some 1,488,300 students took the Scholastic Aptitude Test and Achievement Tests, 166,000 took the Graduate Management Admission Test, 314,000 took the Graduate Records Examination, and 128,000 took the Law School Admissions Test—all tests made by ETS. (ETS 1978 Annual Report.) In 1976-77, 923,241 students took the American College Testing's ACT Assessment, and 60,000 took ACT's Medical Colleges Admissions Test. (ACT: General Information, 1977.)

If students are provided with the basic information required by this Act, these tests will loose much of their mystery. Students will have a better understanding of what the tests can and cannot predict, how their score is determined, if preparatory courses are a worthwhile expense, and whether testing fees levels are proper.

In addition to the question of what the individual student is to be told, there is the larger question of what the public should be told about the limitations of standardized tests, most notably problems with predictability and bias.

One problem with predictability is the lack of preciseness of the test score. For example, if a student received a 600 on his or her SAT (with possible scores ranging from 200 to 800 points), this only means that there is a two-in-three chance that the student's "true-score"—the score the student would receive if all external factors, such as luck in guessing, could be eliminated—would fall somewhere between 570 and 630 points. One ETS booklet revealed that a 72 point difference on the SAT math section (and 66 points on the SAT verbal section) is so statistically insignificant that "it cannot be taken seriously." For the LSAT, a difference of 67 points cannot be taken seriously. Yet 67 points can be the difference of one student being admitted or not to most law schools.

In addition, there is a one-in-three chance that that a student's score will vary up to 60 points over or under his/her "true score." There is a 5% chance that the variance may be even greater than this—ranging more than 60 points over or under his/her "true score." That means that a student whose "true score" is 550, may receive a score over 610 or under 490. Since 1.5 million students take the SAT each year, 75,000 students would fall into this category.

A recent UC Berkeley study by Dr. Harrison Gough also questioned the predictability of test scores. He found that high school students with high grade point averages but low SAT scores did as well as those with high SATs—in fact "they were more apt to take the BA degree (within) four years."

Most important, the tests cannot measure creativity, judgement, experience, or idealism—characteristics which should figure into any school's admissions formula. As many colleges rely heavily on standardized tests to make admissions decisions (to the exclusion of all other factors, except grades), these traits rarely enter admissions decisions.

Testing expert and mathematician Benesh Hoffman, in his book, Tyranny of Testing, also criticized the limited value of multiple choice tests. He contended that the multiple choice tests, such as the SAT, LSAT, GRE, and GMAT, "reward superficiality, ignore creativity, and penalize the person with a subtle, probing mind." Professor Plotkin, a testing expert, has stated, "God help the kid who stops to muse over a test question, or who sees a sublety. And God help him if he gets an idea while he's taking the test . . . These tests look for convergent thinking, not divergent thinking."

There has also been tremendous concern regarding test bias. While the racial, sexual, regional, and ethnic bias of these tests is still hotly debated, it is clear that there is a positive and progressive association between test scores and family income. A 1977 study by the California Postsecondary Education Commission noted that "the average family income for students who earned 650 or more points (out of 800) on the SAT was $26,400. Students in the lower range—below 350 or more (points)—had a mean family income of $14,500."

A 1975 UC Berkeley study supported this conclusion. Students whose family income was below $6,000 scored 92 points lower than a student whose family income was $30,000.

Even some of the data of ACT suggests this is true. In their recent publication, College Student Profiles, Norms for the ACT Assessement (1978–79 Edition), ACT compares the test score of students from varying economic backgrounds. The mean scores for students (on a scoring scale of 1 to 36) is as follows:

Annual family income:	*Students' mean score*
Under $3,000	13.5
$3,000 to $5,999	15.2
$$6,000 to$7,499	16.9
$7,500 to $8,999	17.6
$9,000 to $11,999	18.6
$12,000 to $14,999	19.4
$15,000 to $19,999	19.9
Over $20,000	20.6

This chart indicates a wide disparity between students from different economic backgrounds. The student from the highest economic background will average 7 points higher—or 20 percent of the scoring scale—better than the student from the lowest economic background.

The same publication by ACT suggests some trends in regional bias. The ACT score for students enrolled in four year colleges in the following regions of the United States is as follows:

Region where student is enrolled in college:	*Students' mean score*
Midwest	19.5
Rocky Mountains/Great Plains	19.0
West Coast	18.5
East/Northeast	18.0
South/Southeast	16.2
Southwest	15.8

A student from the midwest will average 4 points—or 11% of the scoring scale better than a student from the southwest.

The same publication also notes that male students generally have higher scores on the ACT assessment than do female students, although women generally do better on the English part of the test.

Questions of predictability and bias need to be discussed. As this Act will inspire a healthy intellectual debate on testing, the public can become better informed as well. In addition, as independent researchers analyze test questions and answers, their research may provide the basis for making these examinations fairer.

I would like to now discuss some of the more important provisions of this Act.

SECTION 3: INFORMATION TO TEST SUBJECTS AND POSTSECONDARY EDUCATION INSTITUTIONS

This section would require that basic information be provided to students taking the test and that information be provided in clear and understandable language. This information is necessary to clear understanding of the tests, including: the purpose of the test; the subject matter of the test and the knowledge to be tested; predictive value of the test; margin of error; explanation of the scoring scale; privacy rights of students; date students will receive their scores, special services for handicapped students; ability of test preparation courses to improve students' scores; how the student's score will be reported; and relationship of students' scores and their economic background.

Despite the importance of these tests on the lives of American students, most do not have a clear understanding of what their test scores mean. This is due in part to the limited information which prospective test takers receive regarding a test and their rights as consumers of test services.

It is true that some of the organizations sponsoring the tests have been extremely responsive in this area. The brand new descriptive booklet for the GRE provides much of the information required by this section, presumably because of the pressure of legislation in New York and California. However, if this pressure no longer exists, will the testing companies return to their old ways?

Often colleges and universities are told some of the information required by this section while students are not. In his testimony yesterday, Congressman Weiss mentioned that students applying to law schools, who had taken the LSAT more than once, had their testing transcripts tagged if the failed to indicate that they had taken the LSAT previously. The students were never told that this information would be sent to law schools to which they were applying.

Lastly, this section requires the testing agencies to immediately notify the student and college, to which he or she is applying, of any reporting delay of a student's score. Although students' scores have been delayed for a number of very legitimate reasons, some students have been denied admission because an excessive delay caused them to miss a college's application deadline. This provision would ensure that both the student and college would be aware of the delay and that the college could take this into account if the delay was excessive.

SECTION 4: REPORTS AND STATISTICAL DATA

This section requires release to the Commissioner of studies, evaluation, or statistical reports for which the testing agency prepares or supplies data. A provision is included to protect the confidentiality of the individual student when this information is made public.

The largest body of research on test validity, bias, and predictability on the tests are produced by the test agencies themselves. This research serves as the basis of improvement of the tests and testing procedures, weighting of tests in college admissions formulas, and development of new services to be offered to students and universities. Clearly, this information should be available to the public and to the scrutiny of independent researchers to foster a better understanding of the tests and the proper role they should play in college admissions decision.

The testing agencies argued against this provision during the hearings on the New York and California testing legislation. They claimed that test research is routinely published and available to the public, academic, and research libraries, and to any individual who may request it.

This simply not true. Many important studies are not disclosed by the testing agencies. Three examples of studies which witnesses at the New York testing hearings indicated were withheld by ETS include:

(1) statistical reports known as "Test Analysis" equating procedures (ETS refused to disclose such data to a researcher who suggested that ETS made a serious mistake in equating the Multi-State Bar Examination);

(2) a major ETS study on "Cultural Bias in Testing: Challenge and Response";

(3) a study on law school's use of the LSAT.

The late Dr. Oscar K. Buros, publisher of the Mental Measurements Yearbook, winner of the ETS award for Distinguished Service to Measurement, and member of the testing fraternity who frequently praised ETS tests, discussed the issue at a March 1977 speech at the University of Iowa. Buros thanked ETS and ACT for making copies of secure tests available to his reviewers, but criticized the testing firms for withholding important data about test characteristics. "The information available to permit an adequate assessment to be made of these secure tests is quite unsatisfactory," Dr. Buros said. "Although our reviewers received some in-house material, which is not available to other educators and psychologists, even this material is inadequate."

Essentially, this section would ensure the free flow of information concerning important testing studies. It would encourage independent research and, in our view, lead to a wider discussion and better understanding of standardized tests.

SECTION 5: PROMOTING A BETTER UNDERSTANDING OF TESTS

This section requires testing agencies to file copies of test questions and answers with the Commissioner of Education only after the test is given and to provide test takers with a copy of their own questions and answers. Test subjects can be charged a small fee to cover the agency's handling costs.

This information shall be made public and copies will be forwarded by the Commissioner to state educational agencies. As this section is essentially the same as the recently passed New York Truth-in-Testing legislation and similar to the California law, the testing agency would simply forward one copy of the questions and answers to the Commissioner and would not have to file the information with each state that has Truth-in-Testing laws. This provision would save the testing agencies time and money.

This section of the bill is important for a number of reasons. First, a basic tenent of fairness dictates that people have the right to review the criteria used to evaluate their ability. Presently, however, students (who must take these tests in order to be admitted to postsecondary educational institutions) cannot see copies of the test questions or their own answers even after the test is given.

Second, standardized admissions tests are major instruments of public concern as they influence decisions about who is admitted to undergraduate, graduate, and professional schools. It is essential that such instruments of public concern be available for public discussion and review.

Third, public access to test contents is essential so that independent researchers can evaluate the tests. Independent review should be welcomed as a way to improve the quality of tests. While researchers are occassionally granted access to test contents now, it is very difficult to obtain this privilege. Furthermore, the conditions of granting access are that the results of this research not be published nor discussed with colleagues outside the testing industry.

Fourth, students who can afford to attend expensive coaching schools already have access to test questions. These schools systematically send staffers in to take tests and memorize the questions for review in the course. Numerous studies have been cited that indicate that coaching can improve a test taker's score. (See Steve Levy, "ETS and the 'Coaching' Cover-up", New Jersey Monthly, March 1979). Granting all students access to past test contents would eliminate part of this severe economic discrimination.

Fifth, contrary to claims made by the testing agencies yesterday, release of questions will not cause problems with "equating", the process that makes test scores comparable. This is because the bill allows the testing agencies to keep secret questions used for equating. As Dr. Walter Haney, Staff Director of the National Consortium on Testing, testified before this Subcommittee yesterday: "Arguments dealing with pretesting and equating of tests are essentially irrelevant to the present version of H.R. 4949, since current provisions allow for maintaining security with respect to items which are used solely for the purposes of pretesting and equating, and which do not contribute directly to an individual test taker's raw score."

Sixth, contrary to testing company claims made at the New York Truth-in-Testing hearings, release of test questions and answers will not result in a large increase in test taker fees. The cost of writing new questions would be minimal according to ETS' internal cost studies (ETS "Activity Analysis", January 31, 1972). Actual costs of test development for the SAT and Achievement tests was $.32 or 5.7 percent of the fee paid by test subjects ($5.75). By contrast, the surplus of income over expenses (i.e. "profit") from these two tests was $.93 or 16 percent of the test subjects' fee. In 1976, less than 3 percent of the income from these two tests was spent on test development.

A breakdown of these costs for these and other tests is as follows:

CHART I: HOW MUCH OF CANDIDATE'S FEE GOES TO TEST DEVELOPMENT?

Test program	College board	LSAT	GRE	ATGSB (now GMAT)
Fee paid by candidate	$5.75	$13.50	$8.00	$10.00
Total cost (per session candidate) to ETS	$4.82	$10.83	$7.17	$9.22
Cost of test development	$.32	$.49	$.62	$.44
Test development as percent of total cost to ETS	6.7	4.5	8.6	4.8

Source: "Activity Analysis", internal ETS study, Jan. 31, 1972.

CHART II: ADMISSION TESTING PROGRAM REVENUE AND EXPENSES

	Revenues	Expenses	Surplus	Percent profit
1974 to 1975	$16,036,276	$12,550,541	$3,485,735	27
1975 to 1976	16,260,652	13,232,474	3,028,178	22
1976 to 1977 [1]	17,640,000	14,099,000	3,541,000	25

[1] Projected.

Source: College board statement of revenues, expenses and fund balance, year ended July 30, 1976, with comparable figures for 1975 and for 1977 current projections.

CHART III: ETS TEST DEVELOPMENT SPENDING BREAKDOWN AND BUDGET FOR COLLEGE BOARD ADMISSIONS TESTING PROGRAM (SAT AND ACHIEVEMENT TESTS)

	1976 actual	1977 budget
Test construction expenses	$291,000	$389,000
Program administration expenses	$10,079,000	$10,864,000
Test construction as percent of program administration expenses	2.9	3.6
Admissions testing program volume (in number of test subjects)[1]	1,737,000	1,430,000
Total construction expenses per candidate[2]	$0.17	$0.27

[1] Test subjects volume figures are from ETS In Fact, an ETS publication.

[2] Expenses per candidate calculated by dividing test construction expenses by candidate volume.

Source: "ETS confidential project operating statement by position—category/project, period ending Mar. 31, 1977," a 66 page computer printout.

Seventh, contrary to claims made by testing agencies at the New York Truth-in-Testing hearings, disclosure of test questions and answers will not result in reduction in the quality of the tests. Test questions and answers will soon be available through legislation passed in New York and California. In addition, this Act will improve the quality of questions writing by subjecting the process to the scrutiny of interested subject matter experts, educators, and the public. In one case where questions of a secure ETS test were disclosed for public scrutiny (the February 1973 multiple-choice Multistate Bar Examination) law professors at the Georgetown Law Center and other Washington, D.C. law schools disagreed on the correct answers to 25 percent of the questions. Following this disclosure, ETS and its client conceded that five of the questions seemed to have more than one correct answer. No tests are known to have been disclosed since.

In general, the quality of now-secret standardized test questions and answers has been a matter of controversy in academic circles for years. Professor Benesh Hoffman of Queens College, the mathematician, physicist, and former Einstein collaborator, has published extensive criticism of the scientific accuracy of the questions and answers which ETS has disclosed in its test booklets. Independent research can verify or refute charges such as this.

Last, contrary to the claims made by the testing agencies at the New York Truth-in-Testing hearings, test makers will not run out of new questions to ask. This claim is inconsistent with the facts of current ETS question-writing procedures as described in the internal ETS "Manual of Information for Test Development Professional Staff". To illustrate this the Manual uses this example of the January and May 1971 Physics Achievement Test: On each version of the Physics Test, 16 questions were drawn from the old item files and 39 questions were newly written by the physics committee of examiners and ETS staff. The Committee met twice and was budgeted to produce 400 new questions (only 56 were actually needed for each test). The Manual notes that surplus questions are produced as a matter of procedure: "The number of items pretested may exceed the number of useable items by as little as 25 or 30 percent or by as much as 300 or 400 percent. In the absence of appropriate rules of thumb, one should plan to pretest about twice as many items as will be needed." Since most questions are already being written anew anyway (with no sign of exhaustion in sight), the surplus questions will provide a cushion. Those found to be satisfactory are already stored for future use.

SECTION 7: TESTING COSTS AND FEES

This section requires testing agencies to file limited amounts of financial information with the Commissioner. It requires far less information than does the California Truth-in-Testing legislation. This section is still necessary because it will provide nationwide information, while California law only applies to that state.

Limited release of income and expenditures will enable the public to better understand the procedures, development, and administration of standardized tests. As I mentioned earlier, students must take one or more standardized tests made by one of two testing companies to be admitted to many colleges. They have no choice as to which test they may take and are not able to shop around for the best price. Thus students are captive consumers of the testing companies and provide much of the companies' income. And since two companies make nearly all the tests covered under this bill, each has little incentive financially to compete with the other.

Questions have been raised as to the large amount of "profit" (i.e. revenue over expenses) of these companies, as well as some of the expenditures made from student fees (such as the luxurious 400 acre ETS campus at Princeton, New Jersey).

In addition, these tests are used to allocate a limited public good—education—which makes the test's manufacturers comparable to public utilities. This increases the public's right to know more about the operations and test development procedures of the limited number of testing companies.

Lastly, limited release of income and expenditures will ensure reasonable compliance with this Act. This section requires disclosure of profits and test question development costs of the testing agencies. If these agencies raise testing fees to compensate for release of test questions and answers of Section 5 of the Act, they will do so under public scrutiny and will have to justify to the public that the increase was necessary.

CONCLUSION

In sum, the Educational Testing Act of 1979 would provide all those concerned with standardized admissions tests with more information about the tests and encourage the responsible use of standardized tests.

I urge your support of this Act.

Thank you.

STATEMENT OF JON HABER, CONSULTANT ON STANDARDIZED TESTING, UNITED STATES STUDENT ASSOCIATION, ACCOMPANIED BY ANDRE BURNETT, DIRECTOR, NATIONAL THIRD WORLD STUDENT COALITION, AND ROGER T. LENNON, ASSISTANT TO THE CHAIRMAN IN MEASUREMENT, HARCOURT BRACE JOVANOVICH, INC.

Mr. **Haber**. I have a rather lengthy statement, which the committee has, and I can submit that and summarize.

Chairman **Perkins**. Without objection, the statement will be received for the record.

Mr. **Haber**. Mr. Chairman and members of the subcommittee, I thank you for the opportunity to appear before you today. My name is Jon Haber and I am representing the United States Student Association (USSA). I was formerly the Codirector of the University of California Student Lobby and was the principal drafter of the California truth-in-testing legislation. I am accompanied by Andre Burnett, director of USSA's National Third World Student Coalition. We are speaking on behalf of USSA's 3 million student members attending over 350 colleges and universities.

We join many other organizations in strongly endorsing H.R. 4949—the Educational Testing Act of 1979—including the National Parent Teachers Association (PTA), National Education Association (NEA), NAACP, National Public Interest Research Group (PIRG), Coalition of Independent Colleges and University Students (COPUS), the Americans for Democratic Action (ADA), as well as many statewide student associations.

The Educational Testing Act of 1979 is designed to provide more information about standardized tests and promote a wider understanding of these tests amongst students, educators, parents, admissions officials, and the public in general.

Equally important is what this act does not do. It does not regulate test contents nor testing procedures of the testing companies, nor any aspect of their admissions testing program. Neither does it specify how colleges and universities shall use standardized tests in their admissions programs. This act is simply disclosure legislation—no more and no less.

I would take exception to the testimony of Mr. Ruud, who claimed that the motivation behind this bill is that many people want to eliminate standardized testing. That is not our intention at all. We simply believe that tests measure a narrow range of characteristics and are therefore subject to misuse and misunderstanding. Through the information provided them by H.R. 4949, students and the public will gain a better understanding of the nature, purposes, uses, and limitations of standardized admissions tests. We also believe that this act will insure that tests are used properly—a goal ETS and other testing agencies supported before this subcommittee yesterday.

Lastly, we have little to say about H.R. 3564—Gibbons. While we applaud the intent of the legislation, if passed, it would have no impact on the tests, and would not improve or change the amount of information presently provided to students. It essentially codifies the testing companies' current practices.

American colleges and universities rely heavily on standardized tests, often using them as the critical factor in deciding which students to admit to their undergraduate, graduate, and professional programs. Although not covered by this legislation, standardized tests also play an extensive role in determining which students will qualify for financial aid, which students will be awarded advanced placement credit, and ultimately, which students can be professionally licensed after their college training.

I have been involved with the question of standardized testing for 3 years. I have spoken and met with thousands of students about tests and at this point will share some of their concerns with you.

Students feel that they are victims, not consumers, of standardized tests. They have little choice but to take the test—their college or university has made it a mandatory component of the admissions process.

Preparing for and taking the test is extremely stressful. Not only are the tests taken during crucial periods in students' lives, but the tests are often the critical factor in college admissions decisions.

While students do not know the test's margin of error, predictive value, or bias, students are well aware of one fact—their score may mean the difference between being accepted or rejected at most colleges. As a result, many students turn to expensive coaching schools or coaching books in hopes of better test scores.

Students are frustrated at being unable to see their questions and answers. They can never know which questions they answered correctly, or in which subjects they need more work. Perhaps they misnumbered their answer sheet or the computer has made another scoring error.

Students are often resentful when they receive their scores. They often feel that the test did not really reflect their ability. Many feel that essay tests would have been fairer indicators of their talents.

Students believe that the tests are overpriced. A student applying to seven law schools must spend as much as $35 simply to take the test and have his/her score and grades mailed out. Additionally, this student might pay as much as $140 in school application fees, not to mention the costs of coaching books or coaching schools to prepare for the test.

I am not speaking of a handful of students. Rather, many students feel this. In 1977-78, some 1,488,300 students took the Scholastic Aptitude Test and Achievement Tests; 166,000 took the Graduate Management Admission Test; 314,000 took the Graduate Records Examination; and 128,000 took the Law School Admissions Test—all tests made by ETS. [ETS 1978 Annual Report.] In 1976-77, 923,241 students took the American College Testing's ACT Assessment, and 60,000 took ACT's Medical Colleges Admissions Test. [ACT: General Information, 1977.]

It is our belief if students are provided with the basic information required by this act, these tests will lose much of their mystery. Students will have a better understanding of what the tests can and cannot predict; how their score is determined; if preparatory courses are a worthwhile expense; and whether testing fee levels are proper.

In addition to the question of what the individual student is to be told, there is the larger question of what the public should be told about the limitations of standardized tests, most notably problems with predictability and bias.

When I was working in California, the University of California was trying to decide how to solve a basic problem. The average freshman could not write as well as the faculty at this time wanted him or her to write. That is a very, very important concern.

So, what happened? The president of the University of California, David Saxon, formed a committee of distinguished faculty members of the university to look at the problem. The committee decided they would use the SAT to solve the problems. Their first proposal would have eliminated one-half the minority students attending the University of California. There was such an uproar, the president sent the faculty back to the drawing board. They came up with another proposal, the Saxon plan, which went before the Board of Regents. An interesting thing happened, people started asking questions. They said, you are trying to improve writing ability of the students. What does the SAT have to do with that? We found that the SAT could not predict writing ability. All it could predict was word association, mathematical reasoning, and so on.

So, we found out that some of the more knowledgeable academic experts in the area did not completely know of the limitations of standardized tests. It is our hope that this bill will provide more information so this does not occur again.

There is a real objection in the lack of preciseness on the standardized tests. For example, if a student received a 600 on his or her SAT—with possible scores ranging from 200 to 800 points—this only means that there is a two-in-three chance that the student's "true-score"—the score the student would receive if all external factors, such as luck in guessing, could be eliminated—would fall somewhere between 570 and 630 points. One ETS booklet revealed that a 72-point difference on the SAT math section—and 66 points on the SAT verbal section—is so statistically insignificant that it cannot be taken seriously. For the LSAT, a difference of 67 points cannot be taken seriously. Yet 67 points can be the difference of one student being admitted or not to most law schools.

In addition, there is a one-in-three chance that a student's score will vary up to 60 points over or under his/her true score. There is also a 5-percent chance that the variance may be even greater than this—ranging more than 60 points over or under his/her true score. That means that a student whose true score is 550, may receive a score over 610 or under 490. Since 1.5 million students take the SAT each year, 75,000 students would fall into this category.

Most important, the tests cannot measure creativity, judgment, experience, or idealism—characteristics which should figure into any school's admissions formula. As many colleges rely heavily on standardized tests to make admissions decisions—to the exclusion of all other factors, except grades—these traits rarely enter admissions decisions.

Testing expert and mathematician Benesh Hoffman, in his book, Tyranny of Testing, also criticized the limited value of multiple choice tests. He contended that the multiple choice tests, such as the SAT, LSAT, GRE, and GMAT, "reward superficiality, ignore creativity, and penalize the person with a subtle, probing mind."

There are other problems as well, such as the question of validity of the test. While the racial, sexual, regional, and ethnic bias of these tests is still hotly debated, it is clear that there is a positive and progressive association between test scores and family income. A 1977 study by the California Postsecondary Education Commission noted that "the average family income for students who earned 650 or more points—out of 800—on the SAT was $26,400. Students in the lower range—below 350 or more points—had a mean family income of $14,500."

A 1975 UC Berkeley study supported this conclusion. Students whose family income was below $6,000 scored 92 points lower than a student whose family income was $30,000.

Even some of the data of ACT suggests this is true. In their recent publication, College Student Profiles, Norms for the ACT Assessment—1978-79—edition—ACT compares the test score of students from varying economic backgrounds.

The student from the highest economic background will average 7 points higher—or 20 percent of the scoring scale—better than the student from the lowest economic background.

The same publication by ACT suggests some trends in regional bias. A student from the Midwest will average 4 points—or 11 percent of the scoring scale—better than a student from the Southwest.

The same publication also notes that male students generally have higher scores on the ACT assessment than do female students, although women generally do better on the English part of the test.

We have really tried very hard to work with the testing associations. Last August, after the California legislation passed, a number of student leaders met with the College Board. We raised a lot of very tough questions. We were promised responses; we were promised another meeting. We have not seen those responses; we have not seen that meeting, and we have been told there is no meeting scheduled.

In addition, Joel Packer, director of the U.S. Student Association, sent a letter last December to the College Board president requesting that he be able to attend one meeting of the College Board to discuss student participation. It took 3 months to get a response, and the answer was no. The letter came from Robert Kingston; president of the College Board. Let me read you part of this letter:

> Well, the executive committee has now held its late winter meeting; I duly presented to them your interest; and they have, for the present, decided that it would not be appropriate to invite "auditors" generally during the course of their business. I think I have to point out, Joel, that the trustees of an organization like the College Board—like other boards of directors, perhaps—are routinely confronted with extremely heavy agendas, a great deal of sometimes detailed and complex backup material, and must make decisions thoughtfully but expeditiously. This is not a strange process, but I think you'll agree that it's one requiring great concentration—concentration of a different kind from that required in a political forum or a public theatre. This particular sense of responsibility has always led the trustees to discourage attendance by any who are not directly involved in the specific items immediately before them, although they customarily and frequently invite special presentations from individuals professionally and expertly concerned with the issues which are before them. Nobody on the executive committee has any illusions about the special relationship of students to the work of the College Board; but in a carefully reasoned discussion they did conclude that this relationship would not justify the serving of a precedent for attendance during the general conduct of their business.

As of this time, we have observed no meetings. These meetings are secretive and we have been denied permission to attend.

I would now like to discuss certain provisions of this act.

Section 3, which requires information of test subjects and postsecondary institutions, is fairly important because it provides basic information to test subjects, because many test subjects do not have a clear understanding of what their test scores mean.

I want to mention that some of the testing companies have been very responsive in this area. For example, in the GRE booklet, the College Board gives students much of the information required in this bill, but this is the first year they have done it. We think it is as a result of the legislation passed in California and New York. We wonder that if the threat of legislation disappears will the testing companies revert to their old ways?

Section 4 of the bill, simply requires reports and statistical data be submitted to the Commissioner of Education. The largest body of research done on the validity of tests is conducted by the testing companies themselves. This research serves as a basis for improvement of tests and testing procedures, weighting of tests, and development of new services. It clearly is important that independent researchers in the field have a chance to look at tests.

There were arguments made yesterday that much of this research is available to the public. This is not true. I would like to list three examples of research cited in the New York truth-in-testing hearings that are not made available.

(1) Statistical reports known as test analysis equating procedures (ETS refused to disclose such data to a researcher who suggested that ETS made a serious mistake in equating the Multi-State Bar Examination);

(2) A major ETS study on "Cultural Bias in Testing: Challenge and Response";

(3) A study on law schools' use of the LSAT.

Section 5 requires release of test questions and answers.

This section of the bill is important for a number of reasons. First, a basic tenet of fairness dictates that people have the right to review the criteria used to evaluate their ability. Presently, however, students—who must take these tests in order to be admitted to postsecondary educational institutions—cannot see copies of the test questions or their own answers even after the test is given.

Second, standardized admissions tests are major instruments of public policy as they influence decisions about who is admitted to undergraduate, graduate, and professional schools. It is essential that such instruments of public policy be available for public discussion and review.

Third, public access to test contents is essential so that independent researchers can evaluate the tests. Independent review should be welcomed as a way to improve the quality of tests. While researchers are occasionally granted access to test contents now, it is very difficult to obtain this privilege. Furthermore, the conditions of granting access are that the results of this research not be published nor discussed with colleagues outside the testing industry.

Fourth, students who can afford to attend expensive coaching schools already have access to test questions. These schools systematically send staffers in to take tests and memorize the questions for review in the course. Numerous studies have been cited that indicate that coaching can improve a test taker's score. Granting all students access to past test contents would eliminate part of this severe economic discrimination.

Fifth, contrary to claims made by the testing agencies yesterday, release of questions will not cause problems with equating, the process that makes test scores comparable. This is because the bill allows the testing agencies to keep secret questions used for equating.

Sixth, contrary to testing company claims made at the New York truth-in-testing hearings, release of test questions and answers will not result in a large increase in test taker fees. The cost of writing new questions would be minimal according to ETS' internal cost studies.

Seventh, contrary to claims made by testing agencies at yesterday's hearings, disclosure of test questions and answers will not result in reduction in the quality of the tests. Test questions and answers will soon be available through legislation passed in New York and California. In addition, this act will improve the quality of questions writing by subjecting the process to the scrutiny of interested subject matter experts, educators, and the public.

In one case where questions of a secure ETS test were disclosed for public scrutiny—the February 1973 multiple-choice Multistate Bar Examination—law professors at the Georgetown Law Center and other Washington, D.C., law schools disagreed on the correct answers to 25 percent of the questions. Following this disclosure, ETS and its client conceded that five of the questions seemed to have more than one correct answer. No tests are known to have been disclosed since.

Last, contrary to the claims made by the testing agencies at yesterday's hearings, test makers will not run out of new questions to ask. This claim is inconsistent with the facts of current ETS question-writing procedures as described in the internal ETS "Manual of Information for Test Development Professional Staff." To illustrate this the Manual uses this example of the January and May 1971 Physics Achievement Test. On each version of the physics test, 16 questions were drawn from the old item files and 39 questions were newly written by the physics committee of examiners and ETS staff. The committee met twice and was budgeted to produce 400 new questions; only 56 were actually needed for each test. The manual notes that surplus questions are produced as a matter of procedure: "The number of items pretested may exceed the number of usable items by as little as 25 percent or 30 percent or by as much as 300 percent or 400 percent. In the absence of appropriate rules of thumb, one should plan to pretest about twice as many items as will be needed." Since most questions are already being written anew anyway with no sign of exhaustion in sight, the surplus questions will provide a cushion. Those found to be satisfactory are already stored for future use. So, it is our feeling there will be more than enough test questions and answers to ask.

The last section of the bill, section 7, is an important section as well. Simply, this section requires testing agencies to file limited amounts of financial information with the Commissioner. It requires far less information than does the California truth-in-testing legislation. This section is still necessary because it will provide nationwide information, while California law only applies to that State.

Limited release of income and expenditures will enable the public to better understand the procedures, development, and administration of standardized tests. As I mentioned earlier, students must take one or more standardized tests made by one of two monopolistic testing companies to be admitted to many colleges. They have no choice as to which test they may take and are not able to shop around for the best price. Thus students are captive consumers of the testing companies and provide much of the companies' income. And since two companies make nearly all the tests covered under this bill, each has little incentive financially to compete with the other. Questions have been raised as to the large amount of profit; that is revenue over expenses of these companies as well as some of the expenditures made from student fees—such as the luxurious 400-acre ETS campus at Princeton, N.J.

In addition, these tests are used to allocate a limited public good—education—which makes the test's manufacturers comparable to public utilities. This increases the public's right to know more about the operations and test development procedures of the limited number of testing companies. Testing companies are like public utilities. It is ironic that we spend a lot of time regulating the cost of electricity, but we have yet to insure in most areas that adequate information is provided to test subjects who must take these standardized tests in order to go to colleges and universities.

Mr. Weiss [presiding]. Dr. Lennon, you are our next witness.

[The prepared statement of Roger Lennon follows:]

STATEMENT OF ROGER T. LENNON, ASSISTANT TO THE CHAIRMAN IN MEASUREMENT, HARCOURT BRACE JOVANOVICH, INC.

Dr. LENNON. Thank you, Mr. Chairman, committee members. I shall read an abbreviated version of it and I request the entire document be put into the record.

Mr. WEISS. Without objection.

Dr. LENNON. My name is Roger Lennon and I appear here as representative of the publishing firm, Harcourt Brace Jovanovich, Inc., where I am employed as associate to the chairman, for testing and measurement matters. Harcourt Brace Jovanovich, through its subsidiary, the Psychological Corporation, is, we believe, the largest of the commercial test publishers. A test publisher since 1918, it proffers a comprehensive set of more than 100 tests of achievement, ability, aptitude, interests, personality, and other attributes. Its test publishing interests are preponderantly in elementary and secondary schools, but it also serves private practitioners, industrial and commercial organizations, postsecondary institutions, professional organizations, hospitals, clinics—in short, all types of test users.

The Psychological Corp., founded in 1921 by a group of professional psychologists, engages in the conduct of testing programs for admission to postsecondary institutions and occupations such as those addressed by the legislation here under consideration—such institutions, for example, as schools of pharmacy, nursing, optometry, veterinary medicine, and other health professions. It offers the Miller analogies test program, used as a basis for admission to some graduate school curricula. The corporation also offers testing programs for licensing, certification, and counseling for such diverse groups as occupational therapists, cosmetologists, operating room technicians, neurosurgical nurses, personnel and industrial relations professionals, computer professionals, marketing communicators, et cetera. These latter programs, or some of them, probably come within the purview of the Gibbons bill, though not of the Weiss bill.

From 1948 until last year I was responsible for all of the Harcourt Brace Jovanovich test publishing activity, and from 1970 for all the test publishing activity of the Psychological Corporation. I am president-elect of the National Council on Measurement in Education, the professional organization of persons concerned with test development and use, though I am not appearing here as representative of that body or of any organization other than Harcourt Brace Jovanovich. It is my understanding that the Association of American Publishers, representing most of the commercial test publishers though not Harcourt Brace Jovanovich, is submitting a statement to this committee on the legislation under discussion.

We welcome the opportunity to appear before this committee and to share with the committee our views on the proposed testing legislation. We endorse the purposes which the Gibbons and the Weiss bills seek to achieve: that persons required to take examinations as a condition for entry into postsecondary institutions or occupations be fully informed about the nature of the examinations and their use; that the public have access to appropriate information regarding the development and uses of such tests; and that tests be used in a manner that insures fair treatment and equal rights for all examinees. We have sought in the conduct of the testing programs for which we are responsible to achieve these purposes. I append to this statement information booklets for several of the programs which we conduct, typical of those placed in the hands of subjects in all such programs, in aid of these purposes.

[Booklets retained in subcommittee files.]

Dr. LENNON. At the same time we question (1) whether there has been a substantial showing of need for legislation of this kind; (2) whether the controls and refinements of practice sought by the bills are not better handled through the appropriate professional organizations, such as American Psychological Association, American Educational Research Association, and National Council on Measurement in Education which have, in fact, addressed themselves, through the development of test standards, to many of the issues of test information that these proposed bills look to; (3) whether, in fact, the bills here under consideration will achieve the heightened level of knowledge and understanding on the part of examinees and the general public that are intended; and, finally, (4) whether the specific provisions of the bills are workable and could be complied with by testing agencies without inordinate additions to their costs and thus the charges to examinees.

Turning first to H.R. 3564, the Gibbons Truth in Testing Act of 1979, we respectfully invite the committee's attention to the following points:

Section 6(a). Here, as elsewhere, questions arise as to the locus of responsibility for compliance. In admissions testing programs, a variety of arrangements may prevail among testmaking agencies, receiving institutions, sponsoring organizations and others, with varied divisions of responsibility for program elements among them. These variances in arrangements and thus compliance responsibility, should be addressed in the bills.

Section 6(a)(3). It is not always possible, prior to testing or at the time of administration, to provide experimentally obtained information on the reliabiity of a test, particularly if it is required that new tests be created for each test administration. It is not inappropriate to ask that information on reliability be provided within a reasonable time after the administration of the test.

Section 6(a)(5). It is not clear what is intended by the phrase "related facts."

Section 6(b)(2). This provision seems to require an interpretation of an individual's performance as compared with the performance of other examinees; this is the only meaning that we can attach to "ranked." We find this requirement inconsistent with that of section 6(c), which prescribes grading of tests of knowledge or achievement on the basis of the relative distribution of scores of examinees.

Section 6(b)(3). This calls for reporting of cutoff scores for admission to institutions of higher education or occupations. Not only are these generally unknown to the testmaking agency but, in many cases, the use of cutoff scores is discouraged as conducive toward a mechanical way of utilizing the test information, possibly to the minimization of use of other admissions criteria.

Section 6(c). In calling for different modes of grading tests of knowledge or achievement and tests of aptitude, this section implies a distinction in the drafter's mind between these types of tests that is hard to reconcile with the treatment of these terms in the professional literature.

Given appropriate handling of drafting infelicities such as those mentioned above, we would not find compliance with H.R. 3564 to require great departure from our current practice. Nevertheless, we advise against its enactment. Compliance would entail additions to our costs which would, of course, be reflected in the costs ultimately borne by examinees. We do not believe that the interests of examinees, or of the public generally, would be appreciably served should this legislation be enacted. We urge the committee to evaluate carefully the evidence of the extent and seriousness of the needs supposedly being addressed by this bill, as well as Weiss.

We respectfully raise one additional point in connection with H.R. 3564, namely, its proposed title, the Truth in Testing Act of 1979. This title seems to us a too facile borrowing from truth-in-advertising, truth-in-packaging, et cetera, legislation, clearly, and perhaps intentionally, implies that heretofore there has been an absence of truth, or perhaps untruth, in representations about admissions tests. This is an implication to which our organization takes vigorous exception. We have been governed in our test publishing practices of the standards of test description and reporting enunciated by the professional groups I mentioned above. It would seem to us that a less prejudicial and pejorative title for the bill is in order.

Let me now turn to the Weiss bill, the Educational Testing Act of 1979. This bill has many of the same purposes as the Gibbons bill: examinees should be fully informed about the content of admissions tests, the public should have access to information about the tests, and there should be assurance that all examinees are treated with equal fairness—worthy goals, all. In seeking to achieve these goals, however, the Weiss bill goes far beyond the Gibbons bill in its disclosure requirements and, in our opinion, imposes such heavy burdens on test-making agencies as ultimately to be counterproductive: the volume of information to be made available to examinees and others and the costs of providing it are, in our opinion, disproportionate to any manifest need or likely benefit.

In particular, we invite the attention of the committee to the following provisions of the proposed bill as matters of concern:

Section 3(a)(3) calls for information on the correlation between test scores and career success. Such information is almost never available, in large part owing to the extreme difficulty of defining and measuring career success. Moreover, most tests used for admissions purposes purport only to predict success in training programs or, where used for certification and licensure purposes, to measure only the knowledge requirement and none of the other attributes that make for career success. Incidentally, the stipulation that correlations be reported is hard to reconcile with the provision—section 3(a)—that information be provided in easily understandable language; very few candidates would understand the significance of correlations of the type called for.

Section 3(a)(4)(A). The type of reporting called for, namely, the extent to which test scores improve prediction of grade point average, expressed as a percentage, is a virtually meaningless statistic, as well as one that would be scarcely comprehensible to most candidates.

Section 3(a)(4)(B). This calls for a comparison of test performance by major income groups. Most programs do not solicit information about income from candidates, and such inquiry would be distasteful to many examinees; but, even if there were dependable information on the income of examinees or their families, comparisons of test performance by income groups is likely to give rise to far more misunderstanding than clarification of the appropriateness of tests.

Section 3(a)(4)(C) calls for information on the extent to which test preparation courses improve examinees' scores on the average, again expressed as a percentage. Such information will rarely be available; witness the difficulty the Federal Trade Commission had in locating data on the effectiveness of SAT coaching programs. Again, the stipulated reporting mode—percentage improvement—is statistically inappropriate.

Section 3(a)(5) calls for information on the manner in which test scores will be used, alone or with other information, to predict future academic performance. In many cases the test agency has no way of knowing how receiving institutions use test scores, what weight they are assigned in the institution's decisionmaking process, what other variables are combined with them in making admission decisions, et cetera.

In a word we believe that many of the reporting requirements with respect to information to be provided to examinees prior to test taking are unrealistic and impractical. While completely sympathetic to the desire that candidates have a better understanding of the nature of the tests they take, we do not believe that such understanding would be greatly advanced for most candidates by the information called for, even if it could be accumulated.

Far more serious, however, and potentially more hurtful, is the requirement in section 5 that copies of the tests themselves be released both for inspection by any candidate so desiring, and for deposit with educational agencies. The destruction of security of a test, consequent upon making it available to all examinees, greatly complicates the task of providing an instrument at the time of each testing that will be equally fair to all candidates, in the sense of equally familiar or unfamiliar to all candidates. We do not advance sheer increase in difficulty and expense to testmaking agencies as per se basis for nonenactment. We do, however, urge the committee to attend most seriously to the effects on costs to examinees and availability of examining services of the required compliance, and to balance the putative benefits to some—and probably few—candidates against an increase in costs to all.

We urge the committee to consider whether there may not be ways of achieving the goals of these bills to a reasonable degree without the kind of total, promiscuous disclosure of the tests themselves here contemplated.

For example, could there not be created in the Office of the Commissioner of Education a test-monitoring agency with which copies would be filed, which would arrange upon request of examinees for an impartial review and analysis of any test, but under conditions that safeguard the security of the test itself?

Similarly, would it not be fairly easy to provide for a mechanism that would insure independent checking on the accuracy of scoring of a candidate's test short of making the test questions and answer key available to all upon request?

We are aware of and sympathetic to legitimate consumer concerns in this context. Tests are important for admission to institutions of higher education, and thus potentially of crucial significance to an individual's career; and that an individual may feel entitled to see the test questions, study them at leisure, find out where he made mistakes, and so on, is eminently understandable.

Yet it is true that with respect to the programs that we have conducted over many years, reaching thousands upon thousands of examinees, there have been very, very few requests from examinees that they be permitted to examine copies of their test booklets, and it is hard for us to see how their fortunes have been hampered by virtue of their not having had access to the tests.

We judge that candidates in these admission testing programs understand the advantages of test security, which is the only way of insuring that no candidate have the advantage of familiarity with some or all of the test content. To maintain test security in the face of the disclosure provisions of the Weiss bill would create substantial and unmistakable added costs for the testmaking agency and thus the examinee, for dubious benefits.

Equally ill advised, in our opinion, are the provisions of section 7 with respect to the reporting of financial information and volume of activity for all admissions testing programs, purportedly to "ensure that tests are being offered at a reasonable cost to test subjects."

In our system of government, the Federal or State Governments should never become involved in determining competitive prices except in those instances where there is a nationwide consensus to combat inflation, to penalize acts of monopoly or price fixing that are contrary to law, to meet a national emergency that applies to all industry, and to regulate those industries which use public property such as roads, tracks, and so on.

There has been no showing, as far as we are aware, that charges to examinees have been unfair. Tests making and test publishing is a relatively small business in total. In our judgment, the price of tests and answer sheets and scoring cannot be reduced by means of Federal or State examination of publishers' financial records and monitoring of their compliance with the provisions of either of these bills.

In the end, should such procedures be undertaken, publishers will employ fewer people, government will spend considerable money, and testing will not be for the consumer a single dollar cheaper, if one considers that test makers and test takers are both taxpayers, whereas, examinees could be severely handicapped if publishers find that test publishing is too expensive to continue under the proposed procedures.

In a word, the test and financial disclosure provisions seem to us discriminatory, possibly illegal, and certainly unworkable.

For the reasons enumerated, Harcourt Brace Jovanovich strongly urges that the two bills here considered not be approved by this committee.

As an organization with "proprietory interests" in admission programs, we see threat to those interests in the Gibbon's bill, and serious harm in the Weiss' bill.

Please let me reiterate that even if what is proposed in these bills proves to be constitutionally correct and popularly sought, we believe that it may be impossible for us as a major test publisher to conduct our business under the constraints and costs imposed. This is not an idle threat.

Rather than break the law, should one or both of these bills be passed, it is possible that we would gradually withdraw from a business in which we have invested more than $40 million over a period of six decades and at present maintain with a staff of more than 23 Ph. D's and 209 other employees which we believe would not be in the interest of candidates in the sense of reducing the available supply of purveyors of materials and services for such programs.

I thank you, Mr. Chairman and committee members.

Mr. WEISS. Thank you very much, Dr. Lennon.

I am going to ask two questions which relate to drafting, and let Mr. Goodling ask questions of the panel.

Mr. Ruud, I have tried to confirm the area of concern that you had expressed about some of the drafting, in regard to the availability or the return to a test taker of copies of the tests and the answers.

I look on page 7 of the bill. It is possible that they had not used the word "copy" but the final draft does. That has been corrected and, Dr. Lennon, with regard to section 3(A)(5) where you express some concern about calling for information on the manner in which test scores will be used alone, or where other information to predict future academic performance, and you say in many cases the test agency has no way of knowing how the receiving institutions use the test scores.

In 3(a)(5) there is a provision which says:

"And the manner, if any, the test agency will use the test scores"; so I think that that concern probably has been anticipated and responded to.

Mr. Goodling?

Mr. GOODLING. There are so many people and so many questions to be asked, and I am 8 minutes late for the next meeting.

Mr. WEISS. Feel free; take as much time as you like.

Mr. GOODLING. I could hear you while I was on the telephone out in the lobby.

I could hear most of your testimony loud and clear, so I am aware of what you were saying. In addition I have read a great deal of the testimony.

First of all, Mr. Herndon, contrary to your expeditious hopes, I would hope that we go slowly in this whole business of passing this legislation, either pieces of legislation. We have great opportunity now to see what happens in New York. They don't know whether they have created a monster or something good and holy, and we can sit back and wait and benefit from their mistakes.

Also, I think National Science Academy, and other organizations are currently looking into this whole testing area. But, let me ask you a couple of questions.

I notice that you supported, and I am sorry to notice that you and AFP appear to be together on these particular issues, you are supporting this legislation. As I understand it you have a real opposition to standardized testing, period.

Am I not correct in saying that?

Did you not take such a stand?

Mr. Herndon. The NEA called for a moratorium on the use of standardized tests that many of the kinds of questions that were addressed in this legislation were answered.

We have opposed many contemporary uses of normal referenced standardized tests, but not the whole concept of standardized tests.

Mr. Goodling. Knowing that teachers feel very threatened with standardized tests, I read somewhere you were in a different position.

Mr. Herndon. Our position has been pushed and distorted to extremities by a number of people that would like to make it look extremely absurd.

Mr. Goodling. Do you believe all standardized tests should be graded and returned to the test taker. This is contrary to many teachers that I worked with who didn't return tests or who returned them and allowed students to look at them a second or two and then they took them back for some reason or another.

Do you believe all standardized tests should be returned to the taker?

Mr. Herndon. That answer has to be a function of the use of the test, what it was intended to do, and the decisions that are going to relate to the scores.

Mr. Goodling. Much of what I have heard in relationship to the legislation is to do that kind of thing.

Mr. Herndon. I believe in regards to the use of the SAT for college admissions or the use of the LSTA for law school admission, the use of the graduate record exam for graduate school, the use of many of these examinations for entry into professional schools, that that is a very reasonable provision, that it ought to be done.

Mr. Goodling. Do you believe that this is the sole way of getting into schools, et cetera, in most instances?

Mr. Herndon. I cannot answer for most instances. There are a number of institutions where we believe the scholastic aptitude test is almost the sole determinant of admission.

Mr. Goodling. End of comment?

Mr. Herndon. Yes, sir.

Mr. Goodling. I can only say it's contrary to my experience, so it must have changed considerably in the last 5 years, because it certainly was contrary to my 20-some years experience prior to that.

Let me ask you one or two other questions.

Yesterday I tried on numerous occasions to find out, and I didn't get an answer to my questions, whether these two pieces of legislation we are talking about at the present time will in any way help the youngster from the low-income family, the minority youngster who we are told from time to time, and I have documented that in many instances does more poorly in these types of examinations.

Do you see anything in these two pieces of legislation which is in some way going to help those youngsters?

Mr. **HERNDON**. I think there is a potential for help. Whether the help will automatically follow is prospective.

The Weiss bill will generate for once the data that are necessary to make accurate decisions and accurate conclusions as to how these tests and the decisions that relate to these tests actually affect those from low-income groups.

Now, what the Congress or the educational institutions choose to do with these indications is another matter.

We would hope that they would respond constructively.

Mr. **GOODLING**. Yesterday I heard on numerous occasions that it was going to do something about the business of the more wealthy student who is able to for preparation for taking the examinations. I am wondering, for instance, about handing out these tests after they have been scored to the people who took them. What is that going to do to generate help to the low-income person?

He is still going to have to have someone then, I suppose, to tutor and prep him in order to take the next test. On the other hand, it seems to me it works against him, because it gives even a great opportunity, I would think, for tutoring purposes.

Mr. **HERNDON**. It's very safe to point out that the statute as proposed does not offer copies of examinations to people prior to the time they take the examination, and it would be available to researchers.

Mr. **GOODLING**. I am talking about the second and third time you take the test, does it in fact give it an advantage or a disadvantage to the low-income person?

Mr. **HERNDON**. This bill as written does neither of those. It generates data that allow us to make decisions as to the real effect.

Mr. **GOODLING**. It does permit the student to have the exam after he has taken it.

Mr. **HERNDON**. Yes, sir, but I think it most unlikely the producers and purveyors of the examinations are going to be rendering the identical examination to the same student.

Mr. **GOODLING**. Not identical, but how many times can you change the same routine?

Mr. **HERNDON**. The factor that you are describing is in existence at the moment. The student scores do go up as they take these examinations more times.

Mr. **GOODLING**. How will this legislation improve the possibility of low income or minority students doing better on these tests?

Mr. **HERNDON**. I think that the statute as written does not directly address the question.

It does provide us with data that allow us to make decisions about the actual effect on low-income students.

Mr. **GOODLING**. I think we have all of that data already. But beyond that, it does make those tests available for a better tutoring program, I would think; and again people who can afford it then would have a better opportunity than they presently have for such tutoring purposes because of the availability of more tests.

It is just an observation. I have not been able to find out.

I was trying yesterday to find out what the push was behind this, what was the motivation, and also if motivation was the low-income and minority students.

Mr. **HERNDON**. It's my belief that, as these data are available, the provision of this statute is implemented we will have better tests and that was one of the quotations I rendered from Mr. Buros.

It is his belief and my belief there are a number of people capable of contributing to the refinement of these institutions, because they couldn't work for one of a handful of corporate interests that promulgate the examinations.

Mr. **WEISS**. In the course of my testimony yesterday, I tried to indicate again, independent researchers in the field who have, on the basis of some analysis and studies they have undertaken, indicated their belief that there are unintentional inherent biases that are built into some of these tests. Making those tests public and allowing independent researchers to review them and pick out the problem areas and eliminate the unintentional biases that have crept in, will improve chances for those it would adversely affect.

Mr. **GOODLING**. I have no problem with your independent researchers' review, but that is different than handing them out to students.

I am thinking in relationship to standardized tests. I never used them other than to determine to some degree what the youngsters' potential is, and then if he is not reaching that or why he is not. Then maybe I can help a student from that standpoint. But, I also know when you get the standardized tests out, for instance, the first district I worked in as a counselor, they had this crazy idea that you had to homogeneously group students, whatever that meant. So there were standardized tests which of course, threatened teachers.

They feel if students don't do well on a standardized test it is because of the teacher. So, I would find one sixth-grade group who would come in with outstanding test scores. It is because basically the teacher was smart enough to teach for that test.

We had one group that looked outstanding. They weren't any more outstanding than the next group. It was just the teacher was a little smarter and a little more threatened than the other.

Mr. **WEISS**. Some students are able to get the benefit of that kind of pretutoring.

Mr. **GOODLING**. But you are not answering my question as to how does this legislation help the low-income individual. He can have all of those tests that he wants, but he cannot do anything with those tests unless he is tutored.

Mr. **WEISS**. Tutoring is in the eye of the beholder, the tutoree and tutoror, and it seems to me if somebody gets a test and goes to his or her teacher or parent or friend and say, "Hey, this is the test I took; these are the results that I got back; what am I doing wrong?"

Mr. **GOODLING**. Mr. Haber, is it?

Mr. **HABER**. Yes.

Mr. **GOODLING**. I notice you say you represent so many millions of students, and I am always concerned about that. We get all sorts of people in front of us. If they can't get something down on the local level, then they come to us and say, those stupid people on the local level, we can't get them to go along with us since they only represent a minority thinking.

You said you talked with a couple of hundred. I am talking about representing millions.

Have you done any kind of poll?

Mr. **Haber**. Let me say two things:

We have 3 million dues-paying members. I have spoken to many, many students about a wide variety of issues, everything from financial aid to tuition to collective bargaining, and there is no other issue that gets students more excited, more involved than this issue.

I can't quite describe it to you. I could take you to a student conference, if you like.

There is going to be one in 10 days in Massachusetts, and I would be happy to take you there, but there are no other issues which students are more concerned about.

Mr. **Goodling**. Do you have material that you have gotten from these students that one could read, for instance?

Mr. **Haber**. Letters and things like that?

Mr. **Goodling**. Yes; as to what their concerns are.

Mr. **Haber**. Sure.

Mr. **Goodling**. I want to make sure who we are talking about. The NEA and PSEA representatives in my district told me that the teachers are gung-ho for the Department of Education, which became the "Department of Some Education and not some Education."

I couldn't find 15 teachers in my district that would tell me anything other than send money back and keep your cotton-picking fingers off of our business.

Mr. **Haber**. If you like, I would be more than happy to set up a meeting with as many students as you would like.

Mr. **Goodling**. I am taking too much time.

Let me make one or two observations with the gentleman from New York, and I appreciate your coming down, and I appreciate that you have this trial experiment going. We should take time to learn from your errors, or your good points, in my estimation, before we jump into something with both feet.

We do that so often, and then we get burned.

Do you still administer that barbaric testing program from the Dark Ages called regents?

Mr. **LaValle**. The answer is "Yes, we do." The very interesting thing about the regents, when I was a young man, there was always the threat that they were going to do away with the regents, and I guess about 30 years have passed and the rumor has persisted but the regents continue on.

One of the former Commissioners even indicated that formed a task force, and the task force recommended that they eliminate the regents, and the Commissioner did an about-face and we still have those regents.

Mr. **Goodling**. I don't notice in your legislation that you covered them.

Was there any particular reason?

Mr. **LaValle**. Very interesting factor, my colleague here, Senator Halperin, as a matter of fact, in the debate in the Senate on this legislation and Senator Lack indicated that, for instance, we almost had a truth-in-testing in the use of the regents.

If you come to New York State and come to a high school you see every boy and girl toting a little book of former test questions and so forth, and the teachers very vigorously teach to those test questions to prepare them, both in giving them a test awareness and also to try and zero in on the format of specific questions.

The most standard one that we all joke about is as the train leaves from Chicago and one leaves from New York and they are both going at *x* number of miles, when will they meet kind of thing, and to give them that kind of an awareness on the rate and distance question, along with how to take it.

Mr. **Goodling**. Why didn't you cover the SAT's? I don't believe you covered the achievement test either.

Mr. **LaValle**. We had a meeting with the ETS people following our hearing and that meeting lasted several hours, from memory of the 2-hour meeting, a good portion was spent by their people on the achievement tests, that they were very specific kinds of questions.

They use, for instance, the Russian exam. Very few students take that exam, and we felt and on balance that their input in terms of replacement of a highly specialized type of question as given on the achievement test made some sense, so we exempted the achievement test, as we did on two other kinds of tests, the equating questions, and the pretest questions that do not in many of the tests make up the raw score, so we did in our legislation look at certain kinds of tests, for instance, licensing exams, tests that are given, certain kinds of elementary school tests that are given, civil service tests, and we exempted those kinds of tests and we even exempted certain kinds of questions to come under the disclosure provisions.

Mr. **Goodling**. Did you respond to my question?

Why you didn't include regents?

I didn't hear a specific reason as to why you didn't include regents.

Mr. **LaValle**. We are talking about specifically postsecondary admissions to get into college or graduate schools, and so forth.

Mr. **Goodling**. Your bill only deals with postsecondary?

Mr. **LaValle**. Yes.

Mr. **Goodling**. Why did you stop there? Is that as far as you wanted to bite off at that particular time?

I think the most damnable testing programs that go on go on in possibly grades 8 through 12.

Mr. **LaValle**. The thrust of the legislation in terms of standardized testing is beyond or is in the postsecondary area.

Mr. **Goodling**. I consider regents to be very standardized.

Mr. **Burnett**. I would like to make a couple remarks with regard to your question about what this legislation will do for low-income students.

I think that you should be aware that these minority students who are taking these tests nowadays, these people are motivated to go on to college and postgraduate work, and so that this legislation is telling them they can get these tests back and they can study from these tests, and that will stimulate more self-motivation on their part to study on their own for the test as well as to go out and seek counseling, the emphasis to continue on and study more, so next time around they will do better.

With regards to students and how they feel about this legislation, the whole field of testing, the minority students have a great distaste for standardized testing.

Those students with whom I have talked across the country are in favor of this legislation, because now it gives them an opportunity to see exactly what they are paying out for and exactly what their tests mean and how they are being judged by the institutions or whatever, what the test actually means, and seeing what the testing companies are doing with their tests and what they are trying to judge.

When I came in I was very naive; I thought that it was going to be strictly medicine, so I studied all the medical type things and all of these other areas I was not aware of were on the test, and that is the type of thing that students need to do.

Everything is being covered, and this type thing will let them see exactly what is going to be covered so they have a way of studying.

Mr. Goodling. MCAT does have a very extensive tutoring program that you can get your fingers on beforehand.

Mr. Burnett. That is an expensive program as well. I got the book and the book alone costs me $10.95, and it has probably gone up since then.

Mr. Weiss. Senator Halperin, do you have a response?

Mr. Halperin. I have here material which was provided by one of these coaching courses, and I make the following offer, that if the test preparers, although they said under certain circumstances they might make their information available, and you said that by making the test questions, past test questions available, they might in fact be helping the coaching school.

It is my feeling as someone who took one of those courses and knows they have professional test takers who take those tests solely for the purpose of going in and evaluating them and bringing out questions, that the material which I have in front of me in fact will be very close to the questions and in some cases may be the exact questions that were given the ETS, these are for the LSAT's, this particular material, and will allow us to go in and compare, maybe we can get the answer.

Maybe your concern is really not a concern, because I feel that information is already public.

Mr. Goodling. Maybe it was one of those professional test takers that delayed my daughter for about an hour, because one fewer test was turned in than was handed out. She couldn't get out of the room.

Mr. Halperin. I would also like to reflect on one other issue, the opportunity of the individual who is taking the test to review the test.

I bring this up with a little bit of embarrassment, since when I ran for office I also was studying for the bar exam at the same time, and you can't succeed in everything all at once.

I had an opportunity to review my bar exam which was only given to those who were not quite as successful as some others and I found to my amazement, in one sense relief, that the reason I was not so successful is that I had in fact failed to answer a whole series of questions because of the particular way in which these booklets were prepared.

They weren't the normal type of booklet which you turn back by page. They opened up in an accordion fashion and I omitted 20 short answers entirely and I was told that that in fact might have made the difference between my passing and not passing the bar exam.

Mr. **Goodling.** The poor guy you were representing would have gone to the electric chair because you made that mistake.

Mr. **Weiss.** Great exit line, I must say.

Mr. **Halperin.** I have yet to see a case book that opens up with accordion folds.

Mr. **Weiss.** Or a complaint or an indictment.

Mr. Williams?

Mr. **Williams.** Thank you, Mr. Chairman.

I want to commend you and the others who have shown an interest in this matter.

It certainly is a very important issue. Testing has clearly become a major part in the American way of life and the American way of success. Particularly when those tests are given at entry to college level, at that point where a person wants to enter a profession.

I believe the institutions which design, distribute, and give tests really have a greater influence, although momentary, a greater momentary influence on students' lives than does government.

Mr. **Williams.** A far greater influence. If one does not believe that, just ask the student who is preparing for a test or who has failed a test. Ask him at that moment which has the greatest influence on his life, the test booklet or the U.S. Government. He will tell you very, very quickly. So, it is an important matter.

It is important to me for another reason, too, and that is I think that some of the professions in this country have virtually closed shops. I do not know if that is good or bad. The dental profession has it, the medical profession, the ABA, I think George Meany should be envious, and I understand he is envious of the union these people operate. The reason they are so closed is because of admission standards and licensing procedure. Licensing is a whole other matter that someday we will get into.

Having said that, I would like to ask Dr. Graham—I was impressed, Doctor, by the testimony that said in the 1930's or 1940's, dental schools had an attrition perhaps three times as great as you had after you instituted the dental testing program. You had a 20- or 25-percent attrition rate before you put in the dental admissions testing program, then a 7-percent rate. So you obviously kept data on that. Let me ask you to expand on that data and tell me if you know which students were eliminated. I assume the reason the attrition rate was down is because you eliminated some students. Which students? Has the ADA collected that as well?

Dr. **Graham.** No, it has not, not as to who was eliminated, no.

Mr. **Williams.** I would think that would be one of the things any profession would want to know for its own future and for those who are being served by the legal or medical community. That is, if you are going to eliminate perhaps as many as 1,200 people a year, you should want to know which ones you are eliminating.

386

Let me ask another question. Did the ADA or did the dental schools consider other ways to reduce the attrition rate? For example, did they check to find out which courses were causing this attrition rate and then check the relevancy of those courses. Maybe they were courses in French. Maybe dentists ought to know French, I think they should, but maybe that is not relevant to being a good dentist. Did they audit that?

Dr. **Graham**. I do not understand your question.

Mr. **Williams**. There are at least two ways to reduce the attrition rate. One is to eliminate through an entrance test those students who may be less successful in school, and that is what you chose to do. Now, there is another way to accomplish it. That is to change the curriculum, not necessarily diminish the quality of the curriculum, but go in and find those areas where the student is having trouble and see if those courses are relevant to the student being a better lawyer. Do they do that?

Dr. **Graham**. Yes. They would get students in medical school not able to perform the manual work. The students found out after 1 or 2 years in dental school they could not perform the type of manual dexterity. The 1,200 students were not thrown out by the schools, it was voluntary.

Mr. **Williams**. It becomes difficult to know——

Dr. **Graham**. When you go day after day to try to do some very fine work that you cannot get and you fall further and further behind, and you realize the profession that you have chosen will require this fine work—that is what happens. We feel, if we do not provide the student some kind of a task that they can judge for themselves before they spend 4 years in college and another 2 years in dental school——

Mr. **Williams**. Mr. Haber, does the United States Student Association concern itself with attempting to determine the relevancy of the courses which the students are required to take with regard to the profession in which they are trying to proceed?

Mr. **Haber**. We try, but usually those decisions are made by faculty members who are very, very skeptical of being told what to teach by students.

Mr. **Williams**. What procedure do you think should be developed to allow students some input? I am talking about having students' input, but not to determine curriculum.

Mr. **Haber**. One of the things we do, and I can talk as to what has happened at the University of California, on a number of committees there are students who come and speak with the faculty members as to various classes or sit on committees that help make those decisions. We are more fortunate because student participation is much greater in other parts of the country.

Mr. **Williams**. Senator LaValle, there is a great deal of criticism and some of it, I must say, is very well founded, about government today rushing in and trying something before there seems to be an appropriate reason to pass the law or to make whatever attempts it is that is being required. I am wondering, inasmuch as two large States, New York and California, have embarked on this, whether it would not be better to wait until we at the Federal level see what your successes are before we commit a Federal bureaucracy to this procedure.

Mr. **LaValle**. I would think we would want to tie together all 50 States in a unified program. There has been considerable interest throughout the country on this legislation. As a matter of fact, California initially filed legislation very similar to New York's. They lobbied very heavily and amended their law and have a much weaker version than that of New York. They would like to try to move back to where New York is. There are many other States which during the debate contacted us.

You can see that New York, for instance, some of the problems where we have to get to the bottom of the threats by both the medical and dental associations, and so I think that all the students throughout the Nation would be served by the legislation such as Congressman Weiss has introduced.

Mr. **Williams**. I was informed the other day that about 20 percent of the postsecondary students are enrolled in the States of New York and California. So it seems to me if the act was the same in both New York and California, all you need is another State or two and you have in effect required the testing people to meet what would be a national requirement. If you are going to meet it for 30 percent of the students, you may as well meet it for all of them, it is more expensive not to do it.

Mr. **LaValle**. I have heard that as a strategy, that if we can lock up another couple of States we will in essence accomplish this. But there is nothing like legislation that is very specific so that all 50 States will have that coverage and that umbrella which we want.

Mr. **Williams**. One final question for Senator Halperin. Are you an attorney now, sir?

Mr. **Halperin**. I am an attorney in New York and I went down to Florida, 4 days' study, and passed the Florida State.

Mr. **Williams**. In either of those States have you lost anyone to the chair?

Mr. **Halperin**. No.

Mr. **Williams**. Thank you, Mr. Chairman.

Mr. **Weiss**. Mr. Kogovsek.

Mr. **Kogovsek**. No questions, just a commendation to you for submitting this legislation. I hope it passes. If it does not pass, the amount of dialog that has been generated has been important.

I, in deference to Congressman Goodling, I think he was asking the wrong question. I apologize for not being here yesterday. Hopefully the question was asked.

It has been proven standardized testing has been unfair to minorities. I think the question should have been, what have the purveyors of the standardized test done, first of all as Congressman Williams was asking, to find out in what areas the standardized testing has been unfair to minorities? Where have they performed badly and what have the purveyors done to make sure it is rectified? If that question was not answered yesterday by the purveyors, I think at some future time it will be answered.

Mr. **Weiss**. I hope you will bear with us. I have a series of questions I hope a number of you can respond to. Some of the questions are raised by some of the testimony we have had over the last 2 days.

For example, we have been told by a number of those who say that it is not necessary for us to have new legislation, that indeed, all right, a lot of information is provided to students in advance. For example, I think Mr. Ruud gave us the LSAT manual; Dr. Cooper says they have a manual they send out for $3.25 and so on, I think the dental association has an advance booklet.

That being the case, Mr. Haber, perhaps I should address this to you. Why are these advance manuals and the kind of responses that students can get now after taking the exams not sufficient? Why do they not meet the requirements of students?

Mr. **Haber**. First of all, I think in regards to what is being made available to students, each of the test booklets tells students things in slightly different ways and in some instances substantially different ways. For example, the booklet on LSAT has extensive information as to how the student's raw score is transformed into the final score.

I am told some of the descriptions of the other tests are inadequate. Just a short sentence or two. Much of the information provided is now being provided for the first time. It is obviously there because of the passage of legislation in New York and California and concern about testing on a national level.

My concern is that some of this information the companies supply begrudgingly. And if the threat disappears they would be much less willing.

Mr. **Weiss**. Mr. Herndon.

Mr. **Herndon**. Had Senator Halperin not seen his score response sheet he would not have known he failed to pass the bar examination because he failed to answer a number of questions. I think in many of these questions there is real debate over the adequacy of the answer, whether it is the proper answer, whether in fact that is the best answer.

The student would not be conscious of any error of score, error in the scoring process. One might say that is extremely unlikely, that may only happen in the extremely rarest of circumstances. But suppose that rare circumstance is me, or it is you. Or it is my children or your children. It really matters little to me that my child is the only one. I still would like to have the opportunity to examine the material to know that is the case and to pursue my remedy if there is an appropriate remedy available.

Neither will the provision of that handbook make it possible for the academicians, the cyclamatricians, the administrator of instructions to examine internal studies as to validity, reliability, content reliability.

I think there are many deficiencies, if one is comparing a prototype booklet based on past experience, with today's reality of the test. They are just not the same.

Mr. **Weiss**. Let me ask Mr. Ruud and Dr. Graham, what do each of your admissions tests test for? What do you hope to get back?

Mr. **Ruud**. Let me first respond by saying it is not ours constitutionally. It belongs to the law school admissions test.

The law school admissions test is not a test of the American Association of Law Schools, but what it tests for is an ability to try to predict performance in law school. They are designed to measure an ability to read, understand, reason logically using a variety of quantitative material. We have an overall SAT score and a two-digit writing ability score.

I must say, if I may add, I remember a meeting in 1955 at the then Executive Committee for Law School Admission Council, now Congressman Drinan, then Dean Drinan, serving on the committee. We were concerned about the matter of providing widely information about the test. So the suggestion that we are new in concern about the student, at least, is not applicable to us.

Mr. Weiss. Let me see if I understand what you have said, though. The tests that you are involved with are not testing for hard factual information. They are not achievement tests. They are really generalized aptitude tests, testing the capacity of the test taker to be able to handle the curriculum of the law schools.

Mr. Ruud. It obviously has some kind of information base, but it is an information base that would be commonly shared by third-year college students. It is not an achievement test but one of certain generalized abilities.

There have been validation studies, anyone who is willing to read can read them in these research reports that have been going on since the beginning of the test, to constantly check the validity of the test, and the test has not been the same—it is not the same test today that it was in 1948 or 1955, when I first became acquainted. There are new kinds of test items. What shows up under test development in the Law School Admission Council budget is a relatively small item. That item is solely devoted to the development of new kinds of test items. I think folks need to be careful of what kind of conclusions they jump to on these numbers. But constantly, we are at work trying to monitor the quality of test.

Mr. Weiss. You heard the testimony yesterday?

Mr. Ruud. Yes.

Mr. Weiss. So you do not find yourself in the situation that the MCAT was in. If the testimony was valid——

Mr. Ruud. And I have no reason to believe it was not.

Mr. Weiss [continuing]. It is their position there is only a certain number of questions they can come up with, and that is because they test achievement at the first-year college level in a number of scientific areas. You have no concern to come up with an infinite number of questions?

Mr. Ruud. We have concern as to how fast it can be done and the cost of doing it.

I think now this is something you cannot say with great confidence, but it appears that by—as I indicated—by lengthening the test, by perhaps tripling or quadrupling the size of the staff at work producing tests, this kind of test, the kind of test of a general ability, but general ability directed at the kinds of intellectual tasks that law students must be able to perform, I think there is a substantial probability the Law School Council could meet your disclosure requirements at some cost to the test taker; exactly how much, I do not know.

But the leadtime on the production of an individual test item, from what I understand it, is a matter of a year to a year and a half to 2 years.

Mr. Kogovsek. Would you yield, Mr. Chairman? I have to run rather quickly, but since the gentleman is here I would like to ask him, what is the process in most of the States, as far as bar exams are concerned, for correcting them? Who does the actual perusal of the test?

Mr. RUUD. I think there is some difference, but let me describe in general terms, and I am not a bar examiner, but I meet with the National Conference of Bar Examiner people from time to time.

There is the Multistate Bar Exam. The questions that did have problems in the beginning—as you do with any kind of testing program—and I gather now, it is on a good basis, that is machine graded. The individual States are told the number of rights. The individual State then decides what is a passing score.

Now, with respect to the essay, which is more like the typical law school problem-solving examination, I think perhaps California is the model of the best way to go about it. They employ people, mostly law teachers outside the State of California, to draft the test questions and prepare a model answer. Those then are reviewed. That is the function of the Committee of Law Examiners of the State of California. It is an integrated bar, a State agency, as you may want—to review those test questions that have been produced, then select the ones they want given the apppropriate mix of subject. Then they employ graders. If my recollection is correct, each essay question is graded by two examiners. It is an anonymous grading system.

You go into other States, my State, the State of Texas—and I have been up in Washington for 6 years now and a little out of touch—but the examiners themselves produce the questions and do the grading of their questions. The quality of the questions, of course, will vary with the experience and so on.

Mr. KOGOVSEK. Thank you.

Mr. WEISS. Let me just stick to this point. Dr. Graham, would you try to respond to that? What do you expect from a test the students take for admission to the dental schools?

Dr. GRAHAM. There are two domains. One is the cognizant domain, the cognitive, and the other is as to dexterity. We have a science test testing the areas of biology, general chemistry, et cetera, and also reading. It is parallel to the text used in dental schools, and we try to see how fast the student can comprehend the material.

The other is a verbal aptitude test. Through that, we hope to be able to give some kind of indication of how a student will perform in his work when he gets into dental school.

Mr. WEISS. Say on a time basis, what portion of time is allowed for the aptitude test?

Dr. GRAHAM. The aptitude test consists of three booklets. The science examination is usually—I have copies of it here. It is a 90-minute test. There are 100 items, 40 for biology; 30 minutes for the reading examination; reading comprehension is 80 minutes, and the verbal quantitative, 60 minutes.

The other part of the examination, which is a spatial perception examination, which includes manual dexterity, there are two examinations in that area. One of them is to deal with two-dimensional relationships. That is 30 minutes in length. The other is three-dimensional, and it is also 30 minutes in length.

Mr. WEISS. Except for the spatial relationship test, which you indicated in your testimony is so limited that you really do not have an infinite number of those to create, as to the other parts of the test, do you know of any reason why there would be a limit on the number of questions that qualify, and experts in the field that would meet your requirements?

Dr. **Graham**. With regard to the DAT, it would be expensive, but it could be done.

Mr. **Weiss**. OK.

Dr. **Graham**. With regard to the National Board examinations, now their subject matter in that area is much more limited and there are a limited amount of items included.

How many ways can you ask a person to fill in a blank? There is only one way to do that.

Mr. **Weiss**. Certification or admission into the practice of dentistry?

Dr. **Graham**. The admissions test on its facts part is not limited. I should explain that a little bit more carefully.

I think we operate on a different philosophical principle. In the MCAT principles that they have specified, the objection or the manner of the tests as to what will be particularly required to get into medical school, we do not accept that philosophy.

We feel that the best performance, the best indication of future performance is past performance, and we want to see how well a student performs on the whole domain of what he learned in biology or chemistry, or whatever else he took in college, so we do not try to limit ourselves to what is required for dentistry and we want to test what has generally been taught across the country. And we feel because we have got a much broader domain, there is a distinction.

I am not contradicting what the other people said——

Mr. **Weiss**. Without passing judgment on which is the best approach, it is fair to say that the MCAT people themselves are relatively new in that different approach as they testified yesterday. And, indeed, after they get through reviewing and analyzing what they have done, they may decide that is not the approach that they like here.

Dr. **Graham**. We have very close contact with the dental schools as we operate this for them, and we did poll them about 6 or 7 years ago and took that kind of thing before them. And unanimously they chose the philosophy that we had.

It wasn't the American Dental Association, but it was the dental students and dental educators themselves who made that decision to keep that kind of philosophy.

Mr. **Weiss**. Senator LaValle, I think he indicated in his testimony that your association did not take the occasion to bring to the New York State Senate's attention the concerns or objections of the legislation that was pending in New York.

Am I correctly stating that, Senator LaValle? I wonder if you have any information, Dr. Graham, as to why that would have been so, as your association's reaction was so quick and strong. I think that is an understatement, once the legislation was adopted.

Dr. **Graham**. I was not with the American Dental Association then. I spent 1 year at the University of Illinois. I was there for 7 years before that, and I am back with the association now.

Mr. **Weiss**. They have missed you.

Dr. **Graham**. If I were there, I would have made vigorous protest. We were informed by various people that the legislation was not going to pass.

Mr. **Weiss**. See, a dangerous assumption. It is very difficult for someone in Chicago to follow all of the legislations that are going on for the individual States really, let me tell you, and we have a difficult time doing that.

You think it is easier when you have a Federal piece of legislation to review; right?

Dr. **Graham**. The third thing I would like to say is that we did make written recommendations in there, as I understand it, and this is my understanding of it. And we did contact the dental schools and our dental societies in New York and asked them to write letters, to send representation to government hearings before we did sign that.

Mr. **Weiss**. Before he signed the bill, but by that time, as you know, the legislative process is completed.

Dr. **Graham**. We did submit, I believe, written testimony about that. I believe we did. I wasn't at the American Dental Association at that time, but it is my understanding.

Mr. **Weiss**. You know the purpose of hearings, this hearing, all legislative hearings no matter where they are held, is to get input in sufficient time so that if there are valid concerns or objections expressed, the legislation can in fact be written in its final form to take note of legitimate concerns.

In a sense you are unfair to your own organization—I don't mean you individually—if that organization foregoes that opportunity because when you come in after the fact it is much more difficult.

Mr. **Ruud**. I just wanted to make it clear when I responded to you about whether an indefinite quality of LSAT test items could be written you really need to address that question to the Law School Admissions Council.

I am giving you my impression only and not making a representation.

Mr. **Weiss**. Will you remind me before we finish the session today to ask you to go through again what the relationship of the various organizations that you are identified with is. That is the hardest I have had in these 2 days of hearing, following who is involved with whom and in what fashion.

Mr. **Ruud**. I will do it quickly right now. I am employed as the executive director of the Association of American Law Schools. We are an association of 136 schools. We have no propriety economic interest in the law school admissions test. The law school admissions test is owned by the Law School Admissions Council, incorporated under the education laws of the State of New York. It has a contract with a separate entity also incorporated under the State of New York known as Educational Testing Service.

The Law School Admissions Council has just incorporated under the laws of Delaware, Law School Admissions Services, Inc., and it is considering having it perform, and it may perform, quite a bit of the services that are now being performed by Educational Testing Service for LSAC.

Personally, I am a former president of the Law School Admissions Council. I was president when we were incorporated in the State of New York. I have spent, as I said, some 25 years in admissions, and I think I know something at least about the dynamics of it as far as the law school world is concerned.

Mr. WEISS. So that, again, I can relate it to other organizations that appeared before us, are you in essence the equivalent of the College Entrance Examination Board?

Mr. RUUD. Not in any sense at all. The College Entrance owns, as I understand it, TSAT, and it has administered or has a lot of its work done in the administration of that program by a contract with Educational Testing Service. We have no interest at all in the test in any kind of propriety sense.

Mr. WEISS. Do you charge a fee to students who take the examination?

Mr. RUUD. The Law School Admissions Council does. We have nothing to do with the law school admissions program other than our 136 schools of the 168 nationally accredited schools, accredited by the American Bar Association—we are consumers of the test scores.

Our admissions committee uses it, grade point averages and a host of other things to make admissions judgment.

Mr. WEISS. Do you have any input in the creation of those tests? Do you make policy?

Mr. RUUD. Institutionally the Association of American Law Schools attend the meetings of the board of trustees and some of our staff attend meetings of the LSAC's standing and special committees on a liaison basis and their president, their executive director attends meetings of our executive committee and our annual meetings.

We do speak up when they are considering a matter that we think has some special impact on the program of legal education generally.

As a practical operating matter, we tend to leave the labor to the Law School Council. We are interested, we observe it and we are in a sense one of the independent, external forces to which that program is accountable.

Mr. WEISS. But you are not saying, I assume, that the individual law schools or your association in essence turn over the responsibility to determine weight that is to be accorded to particular components of admissions or the final judgment as to whether he is to be admitted to anybody else?

Mr. RUUD. The Law School Admissions Council offers a validity service to the member schools. The validity service is one in which they take the school's experience, the class that was admitted, let us say, in the fall of 1977, LSAT scores, cumulative grade point average, law school average at the end of the first year and then develop through a regression equation formula a statement that says, this formula is the one that on the basis of the experience with the fall, 1977 entering class, assigns the appropriate rates to grade point average and to LSAT score.

Then as part of the Law School Data Assembly Service, which is a centralized transcript analysis service, a school may ask that among the material generated for the school is a predicted first year average, they may use that, the product of the research that has been done at ETS under the LSAC or under some other formula.

They made judgment as to which prediction formula they use or admission index. My school uses a four digit admissions index.

Mr. WEISS. Four year?

Mr. RUUD. Four digit. They use an admission index as a way to rank students. One of the problems we have, unlike the situation that existed when I entered the law school of the University of Minnesota, we now cannot admit everyone who has a probability for success. We have to make choices.

A very important item about an applicant is the probability of the success and where that person is likely to rank in the class. These are only probabilities, they are no more than that, and part of the judgment that an admissions committee makes about an idividual applicant is in a sense a refining on the base of judgmental nonquantitative factors, whether that person has a predicted 75 at my law school which is the bottom B.

Two-thirds of the people will get a score between 70 and 80. One-sixth will get above 80, one-sixth below 70. The question is how to look at everything you can find out about the applicant and figure out, is this applicant going to beat the odds or not beat the odds?

Here is a person who had a very heavy workload during his college career, we are able to arrange some financial aid and reduce the workload, your guess then is this person is going to beat the odds because the grade point average reduces the product.

That is only a part of the admissions judgment. To me there are two other very important judgments an admissions committee makes. One, you are going to admit some folks to your class because they are going to make a very important contribution to the education of the other students.

This was recognized in *Bakke* and recognized by Mr. Justice Powell as a basis for taking into account minority and ethnic background in making the judgment.

I don't know whether you are a reader of Doonesberry but Ms. Caucus, who went to Berkeley, I would admit for that reason. She had the kind of experience to make her a very important member of her class.

Mr. WEISS. I missed that sequence.

Mr. RUUD. That was a while ago before the move from the Post to the Star.

Then a third aim is you select the persons likely to make a very special contribution once in the bar. Our schools have been, among other reasons, interested in minority applicants for this reason. We have a desperate need nationally for more minority lawyers and we are trying to serve that, and I think our record is good and our schools have beaten the numbers.

To suggest that the quantitative data about the minority applicant has kept the minority applicant out is to fly in the face of the findings.

The Law School Admissions Council had a study made of the fall, 1976 entering class. We were very interested in that being made because we wanted in our Bakke brief to be able to tell the court what the consequences to the country would be if they took a view that admissions to professional schools could not take into account racial or ethnic status.

That study showed that if it had not taken that into account, 60 percent of the blacks who attained admission would not have attained admission and 40 percent of the Mexican Americans would not have.

That is to demonstrate the importance of the affirmative action efforts at our member schools.

That is perhaps too long a statement to try to give you an impression of the dynamics of the law school admission process.

Mr. WEISS. I am appreciative. It is an intriguing area. I think that as you get more and more into it, more and more questions really are raised, I am sure not just in our minds, but probably in your mind, as you work with this constantly as to the basic validity of the tests themselves because if those turn out to be problem filled and to be flawed, the whole system, in essence, is built on a very weak foundation.

Mr. RUUD. Charley Runyon, now at the State Department, was then an associate dean at Yale in the early sixties. In the latest report that I have in the bound volume, on page 723 of the third volume of these reports, they have indicated that the tests can be used in the same way with the minority applicants as the nonminority applicants.

It predicts as well. In fact, the last study indicated with respect to blacks it may predict a little better. By that we mean that the bell curve is a little higher and not as flattened out.

Mr. WEISS. One phrase that you used in the course of your response again I find significant because I think even today you may have expressed some caution about the bill using the language, and certainly yesterday, we got into a great big debate over the question of the appropriateness of using the word "success" in the predictive factor built into the legislation.

We got into quite a dialog as to what in fact we were doing by using that word in predicting the success, either academically or in a career. We went to the extent of finding in our files a July 17 release from MCAT where they announced they were withdrawing from the State because of the new legislation in which it was said that the test is designed to help predict how well a student will perform in medical school and later as a practicing physician.

There seemed to be the suggestion then that it is OK to predict performance as a practicing physician, but not success, because what do you mean by success?

You used the term "probability of success". I gather from that, that in fact it is not such a forbidden word after all; that the word "success" is used in predicting future courses of conduct.

Mr. RUUD. I am talking about performance in law school and we have two kinds of predictions: One, a prediction of what the first-year average would be, and then it is possible to also convert that into, let's say, six chances out of eight of getting a passing grade of 65, again using the Texas Law School numbers which would be success, that is as distinguished from failing out.

I guess that is an appropriate use of success. I have the same problem, though, when we use the term success in speaking of the profession. We don't use it. We talk about professional performance, but even that is difficult and that is a study that is still under way if we can find the funding to pull it off.

Mr. **Weiss**. Let me ask Dr. Lennon a question and again this ties into this area of achievement, via aptitude test. Can you give us some enlightenment as to how you view, in the working of your organization, the difference in those tests? Are the terms in fact clearly distinguishable? Do they overlap?

Dr. **Lennon**. I think a fast summary of the professional literature is that what we will call tests of cognitive funding—I apologize for sounding a little grandiose. Any test of cognitive functioning is really a measure of achievement. It measures something that has been learned, some piece of knowledge or skill that has been acquired by the examinee as a result of what has happened to him or to her in that person's life.

Therefore, every test is an achievement test in a sense that it measures something that has been learned or achieved.

The distinction between tests as achievement measures and aptitude measures is not a function of the particular question, that is the skills or knowledge, but rather of the purpose for which the tests are used.

Where the purpose is to look ahead, to forecast probable future performance, we tend to say that is an aptitude measure. In a sense that is different from the man in the street's use of the term. The typical man-in-the-street interpretation is aptitude somehow or other is fairly close to some constitution only or inborn or invariant characteristic. That is not the usage of the measurement of the professional.

Achievement tests in the sense in which we have used that term here today, in connection with the medical admission tests, are a real measure of biological information, clearly an achievement test but in terms of the purpose for which the score is used, that is an aptitude test.

We publish the most widely used test called an aptitude test to high schools. If you look at that test, there are sections in it that to the naked eye look very much like tests that we call achievement tests, spelling, English, arithmetic.

What is the justification for calling it aptitude? It is that its purpose is different from the purposes for which we use tests we call achievement tests. There you want to measure present status, you want to measure how much has been learned over the period of time or you want some diagnostic information about skills, and so on.

So if for whatever reasons it is thought beneficial to maintain in the bill those terms to connote some distinction in the drafter's mind, then I think those intentions ought to be made clear in the definition section because they are not in the present language used in the sense in which now most measurement textbooks or other measuring literature would use.

I would say there is no reason why the terms cannot be used in a special sense in the bill as long as the intended discrimination is made clear.

In the Gibbons bill, for example, I think I understood after his presentation yesterday—certainly better than I had on reading the bill—what it is he has in mind in making the distinction himself between measures of achievement and measures of aptitude.

I would still not be sure that I have caught on to what he is after. All I would say is I would like to see the definitions made clear of which kinds of tests are contemplated where there is a distinction made between them.

Mr. **Weiss**. We now find ourselves in a situation where I have to go down and vote. I still have more questions. I don't know how tight your schedule is. We could just as easily handle it by submitting written questions to all of you and get responses back, and that may perhaps be the most effective way of doing it.

I am very grateful to all of you for your participation. You have managed to raise, I think, many, many more questions than we had anticipated. Thank you very much for your cooperation.

The meeting stands adjourned subject to the call of the Chair.

[Whereupon, at 12:57 p.m., the subcommittee was adjourned, to reconvene subject to the call of the Chair.]

TRUTH IN TESTING ACT OF 1979; THE EDUCATIONAL TESTING ACT OF 1979

MONDAY, SEPTEMBER 10, 1979

House of Representatives,
Subcommittee on Elementary, Secondary,
and Vocational Education,
Committee on Education and Labor,
Washington, D.C.

The subcommittee met, pursuant to notice, at 9:30 a.m., in room 2175, Rayburn House Office Building, Hon. Carl D. Perkins (chairman of the subcommittee) presiding.

Members present: Representatives Perkins, Andrews, Miller, Weiss, Kildee, Kogovsek, Goodling, Erdahl, and Hinson.

Staff present: John F. Jennings, majority counsel; Nancy L. Kober, staff assistant; Jeffrey W. Brennan, staff intern; Richard DiEugenio, minority legislative associate; and Marty LaVor, minority senior legislative associate.

Chairman **Perkins**. The Subcommittee on Elementary, Secondary, and Vocational Education is continuing hearings today on two pieces of legislation dealing with standardized tests.

H.R. 3564, introduced by Congressman Gibbons, would require that certain information about educational and occupational admissions tests be made available to persons before and after taking these tests.

H.R. 4949, introduced by Congressman Weiss, affects educational testing for admission to postsecondary institutions, and mandates disclosure of test questions and correct answers after an examination has been given, upon request of the test taker.

The subcommittee held 2 days of hearings on these bills in July and August. These hearings revealed the complexity and importance of the issues surrounding the two bills. We have invited a panel of distinguished witnesses today in order to more thoroughly explore all the issues raised at the last set of hearings.

I will call the panel, and we will withhold our questions until the entire panel testifies.

Will Mrs. Virginia Sparling, president of the National PTA, come around and take her seat?

Dr. Philip Rever, director, Washington office, American college testing program; Dr. Vito Perrone, dean, school of education, University of North Carolina; Dr. Frank Snyder, McGraw-Hill, representing the Association of American Publishers, and Dr. Paul S. Pottinger, executive director, National Center for the Study of Professions.

We will hear from you first, Mrs. Sparling. We are delighted to welcome you all here.

Go ahead, Mrs. Sparling.

STATEMENTS OF VIRGINIA SPARLING, PRESIDENT, NATIONAL CONGRESS OF PARENTS AND TEACHERS; PHILIP REVER, DIRECTOR, WASHINGTON OFFICE, AMERICAN COLLEGE TESTING PROGRAM, ACCOMPANIED BY RICHARD FERGUSON, VICE PRESIDENT OF ACT'S RESEARCH AND DEVELOPMENT DIVISION; VITO PERRONE, DEAN, SCHOOL OF EDUCATION, UNIVERSITY OF NORTH DAKOTA; FRANK SNYDER, McGRAW-HILL, REPRESENTING ASSOCIATION OF AMERICAN PUBLISHERS, ACCOMPANIED BY ROY H. MILLENSON, AAP, WASHINGTON OFFICE; PAUL S. POTTINGER, EXECUTIVE DIRECTOR, NATIONAL CENTER FOR THE STUDY OF PROFESSIONS

STATEMENT OF VIRGINIA SPARLING, PRESIDENT, NATIONAL CONGRESS OF PARENTS AND TEACHERS

Mrs. **SPARLING.** Mr. Chairman and members of the subcommittee, I wish to thank you for the opportunity to offer testimony on H.R. 4949, The Educational Testing Act of 1979.

I am Virginia Sparling, president of the National Congress of Parents and Teachers, an organization of over 6.5 million members.

The PTA endorses the Educational Testing Act as an important step in providing students, parents, and educators with more information about standardized tests. The PTA's support for this legislation is not to be construed as opposition to testing, but rather viewed as a concern that the public be better informed of the nature, purposes, uses, and limitations of standardized testing.

The PTA feels that the proposed legislation insures the proper use of tests by requiring testing companies to disclose information which will help reduce misunderstandings and avoid misuse.

I would like to comment briefly on the major concerns of parents and teachers that have prompted PTA to endorse this legislation:

(1) Accountability in the testing process: From early childhood through entry into professional schools or jobs, standardized tests make significant judgments about our children. Yet no portion of the public school system is less understood than the facts about standardized tests, their assets, and their limitations.

The National PTA is deeply concerned that testing programs are being authorized without any broad public discussion of the issues. As parents and teachers, we feel it is important to achieve a greater degree of accountability in the testing process yet such a review of public discussion is difficult because of the security that surrounds these tests.

Due to the mask of security, the public has been unable to evaluate dispassionately the claims of testmakers. Information about the tests and the testing process is minimal and is rarely shared after the fact with those most intimately concerned students and their families.

Yet these tests serve as gatekeepers to the future for our children; they develop patterns of expectation by teachers and school systems that often are totally unrelated to a child's true capabilities; and because the power to test is the power to influence curriculum, there is a tendency to reinforce curriculum patterns that respond only to the ultimate goal of looking good on the national tests.

An evaluation of assessment requires a great deal more information than is now available to parents and students. The barrier of security now keeps us at arm's length from the process. That is why we look to legislation like H.R. 4949 as a way of not only accomplishing specific improvements, but as a method of providing us with more information about the testing process and use of the testing results which shape children's lives.

(2) the purpose of testing: To improve children's education:

For years, experts have told us that there are flaws in the present norm referenced standardized testing process but that we must use it because "it is all we have." As parents, we are no longer convinced of that argument.

Testing must be responsive to the major goal of a public school system, improvement of the education of children. We know that many school systems are moving toward testing systems more in keeping with local goals and that children are being tested in every classroom not so much to compare them with other children in distant schools systems, but to determine whether they are learning in accord with the objectives of each local school system, and to determine the effectiveness of the system itself in implementing these goals.

(3) the impact of national sampling on local school antonym: Sampling of a national basis may not provide meaningful comparisons. The approximately 16,000 local school districts produce a variety of local educational objectives, and different timetables at which specific skills are emphasized. Nationwide tests are not adequate substitutes for local assessments.

Unless we are planning to move into a national curriculum, these differences of timing and emphasis will continue to exist. But a child should not be penalized because a local school system has chosen a different sequence than the testmaker. Nor should these important determinations pass from the hands of local school boards who are accountable to their communities to the hands of test producers who are not accountable at all.

(4) is the balance of time use in a classroom changing from "teaching" to "testing"? As public school students return to classes they can expect to take more tests than ever before, which leads to questions such as: Is this an effective use of the professional educator's time?

Is this the most productive use of a child's limited classroom time; and are there ways we could get the answers we need without further disrupting the already fragmented schoolday?

The fifth concern, the PTA is concerned that children be tested on what they have been taught. Without this knowledge we do not know whether the system or the child is responsible when learning does not proceed in an expected rate or pattern.

If there is to be any accountability of local schools and if testing is to be used as a developmental tool so that a teacher can specifically pinpoint the next appropriate steps for each youngster, this is essential.

Parents and teachers are concerned about who participates in test construction and the basis on which these people are chosen. This information would be useful in understanding the way in which tests are constructed.

Knowledge concerning what tests purport to measure is critical so that judgments are confined only to what is measured. Testing companies report they are blameless for the misuses of test results yet there are often conflicting points of view among various school systems about what the test results indicate and what information can be gleaned from the test scores.

If test manufacturers must define the purposes of the tests, parents and teachers will be better able to restrict the use of such data to the purposes for which they are intended; and there is an information void on how standardized tests address the unique testing needs of handicapped and disadvantaged youth.

Our seventh concern is test usage, how tests are used, when and by whom is information desperately needed. There is some concern about a redundancy of testing and a discussion is necessary to determine how selective we have been in accepting a whole range of testing to which we subject children.

Multiple choice or short answer type of questions limit the ability to measure a full range of skills. For example, the ability to write well, to organize and transfer thoughts to paper, requires testing of a written sample, yet for many years we have tested writing skills by multiple choice questions.

Why? Because it was easier to score. We reduce writing to the rote principles of grammar, spelling, and punctuation, and then wondered why children's writing skills deteriorated. We must recognize that methods of testing must be made on criteria other than what the computer can efficiently score.

There is no clear indication of who is responsible when national tests backfire or are misused. At present some problems are caught by random spotting; others are suspected but never verified. Is the testmaker responsible? Is the school responsible? Or is it only the student user?

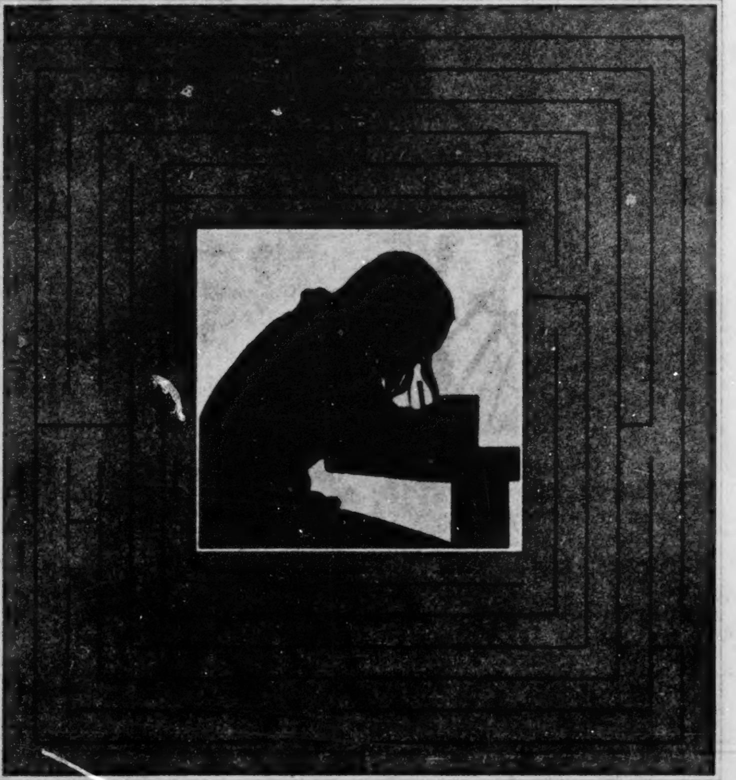
Certainly, tests are only one measure of a person's ability, but because they are so broadly used, the significant questions about their effect and reliability must be thoroughly aired and other potential measurement techniques carefully considered.

I would like to insert into the record the attached excerpt from a recent publication of the National PTA, the Testing Maze. Note pages 16 and 17, a particular excerpt dealing with parents and teachers' concern that we look at children's strengths and weaknesses in a realistic manner, and suggests many ways of assessing a child's growth.

[The information referred to above follows:]

THE TESTING MAZE

AN EVALUATION OF STANDARDIZED TESTING IN AMERICA



THIS PAPER WAS PREPARED AND PUBLISHED BY THE NATIONAL PTA

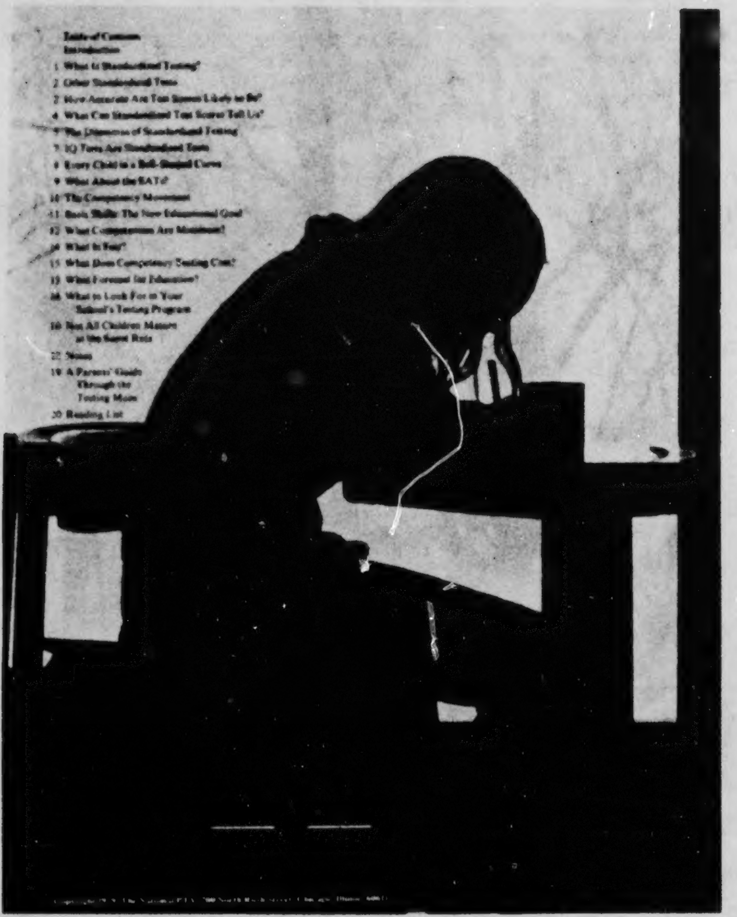


Table of Contents

Introduction
1 What Is Standardized Testing?
2 Other Standardized Tests
3 How Accurate Are Test Scores Likely to Be?
4 What Can Standardized Test Scores Tell Us?
5 The Dilemma of Standardized Testing
7 IQ Tests Are Standardized Tests
8 Every Child in a Bell-Shaped Curve
9 What About the SAT?
10 The Competency Movement
11 Basic Skills: The New Educational Goal
12 What Competencies Are Measured?
14 What Is Fair?
15 What Does Competency Testing Cost?
15 What Forecast for Education?
16 What to Look For in Your School's Testing Program
16 Not All Children Mature at the Same Rate
22 Notes
19 A Parents' Guide Through the Testing Maze
20 Reading List

THE TESTING MAZE

by Dr. Charlotte Ryan

Testing — as evaluation of experience — is a part of all teaching and learning. Teachers need "feedback" from students in order to determine how well they are communicating and to check on student progress in attaining skills and grasping ideas.

Students need testing, too, in order to determine how well they have absorbed new subject matter. Both teachers and students use the results of tests to report progress to parents.

No one test — or group of tests — can evaluate fully a student's work. It is even harder to evaluate a student's ability. Any single test can tell—not what a student "can do"—only what a student *did do* with that particular content at that particular time. Thus schools commonly use several kinds of tests at different times to get a better picture of a student's strengths and needs.

Some qualities such as individual talents and attitudes cannot be measured; they can be identified only with limited accuracy. Recognition and appreciation of these qualities are beyond the realm of tests.

Parents, teachers, and students can learn more from the tests given in their schools if they understand that (1) there are different kinds of tests with different purposes; (2) all tests are limited in what they measure and what they can tell about the student; and (3) misuse of tests can be harmful to students.

Tests have other purposes, of course, beyond providing feedback for teachers and students. Other people besides parents and teachers want an account of student progress —people who want to measure progress by comparisons with other students in other schools.

School administrators, for instance, seek a basis on which to evaluate the curricula and the quality of instruction in their schools — a basis on which they can consider improvements; also a basis on which they can report to school boards and to the public.

People who move into a new community; employers and college people who are critical of high school preparation; legislators and taxpayers who question the use of public funds; researchers in education who want statistical data — all these groups are interested in comparing the effectiveness of one school with another according to some common measures. In satisfying all these interests, schools have come to depend on what is known as "standardized testing."

Standardized testing is now a regular part of school life for 90 percent of American schoolchildren. From kindergarten through high school, a child may take six full "batteries" of achievement tests — maybe more — in addition to other standardized tests, such as those that gauge reading readiness, IQ, and scholastic aptitudes for college.

These tests are sent to be scored by computer in a relatively few centers throughout the United States. Ordinarily, the tests are developed over several years with extensive research and sample testing. They are based on what the test-makers think schools will be teaching as far as ten years ahead.

Because the test items are kept "secure" by contract with the schools that use them, they can be used repeatedly in the same tests or in other combinations. Test-makers say that if new tests had to be developed more frequently, the costs would be two or three times as great. The test items have ordinarily remained in use for 20 years.[1]

Both commercial publishing houses and not-for-profit organizations are among the 45 publishing firms that provide between 200 and 300 million tests each year.

American school systems are estimated to have spent as much as a quarter of a billion dollars last year in the variety of activities related to giving standardized tests to more than 40 million elementary and secondary school children.[2]

This would include the costs of preparing the tests as well as the costs of scoring them by computer. Additional expenditures for textbooks and materials resulting from or relating to tests; the time spent by teachers, staff, and students; the growing investment in coaching students for tests; the costs of certification procedures, as well as college placement and scholarship services dependent on the tests — all of these are thought to add considerably to the funds spent in behalf of standardized testing, which has by now become a major industry built around American schools.

WHAT IS STANDARDIZED TESTING?

How does a "standardized" test differ from the tests that teachers devise and give periodically to students to check on instruction in their own classrooms?

The first difference is that a "standardized" test could be any test which large numbers of students take under similar conditions.

A standardized *norm-referenced achievement* test, which is the type most widely used, is given to thousands or even hundreds of thousands of students in many different schools and school systems throughout the country. The tests do not relate so much to what students may be studying in school at any particular time as they do to general "constructs" — such as reading comprehension, number concepts, or computation.

The second difference is that the scoring of a norm-referenced standardized test is different from that of a regular classroom test. The tests do not represent what a student should accomplish, but rather a range of what students on a given grade level *ordinarily* accomplish.

The tests are set up so they can be scored by computer. That is, a student is given a choice of answers to each question, and a box to mark — which the "scanner" picks up.

The test "items" are tried out with samples of students who represent a cross-section of students to be tested. The results of the sample testing are used to establish "norms" of performance by which other students who take the test can be scored.

The objective of the test-makers is to sort out and rank students according to these norms of performance. Therefore they devise a test in which half the students will score *above* an average point and half *below* the average. The tests are revised until this outcome is achieved. This average point then becomes the "norm" or "grade level" for that group of students. Other students who take the test are scored in relation to that norm.

The test scores are reported in terms of "percentiles." A student who scored at grade level would be reported at the 50th percentile. It is not expected that many students will achieve high scores. For instance, a third grader who answered correctly 33 of 60 items on a certain Stanford Achievement Test in reading would be scored at the 50th percentile: exactly at grade level.

The placing of a student in relation to this norm may not conform very closely to actual classroom performance. Nevertheless, those whose test scores fall under the norm are likely to feel — or be made to feel — a connotation of failure.

OTHER STANDARDIZED TESTS

Some tests ask, simply, "What do you know?"

A norm-referenced test asks, "What do you know on this set of questions in comparison to what other students know?"

There are other kinds of standardized tests which ask, "What do you know compared to what the test-maker knows?"

The well-known *Scholastic Aptitude Test* (SAT), for instance, is a straight measure of performance on fairly stringent tests of verbal and mathematical skills.

The SAT tests college-bound students who choose to take the test — not all do — for likelihood of success in college. It is not norm-referenced.

Criterion-referenced tests are gaining in popularity because they relate directly to certain specific instruction. They measure a student's mastery of specific objectives and thus give feedback on the individual student's strengths and weaknesses.

Domain-referenced tests are similar; they are based on a broader area of study, such as elements of English or math. They might give a longer-term measure of instruction.

Teacher-made criterion-referenced tests have some advantages in being fitted to the work of that particular class. Some say, however, that the tests should be comparable — that is, standardized — if the scores are to be credible. It is said that the instructional objectives — and the scoring — should be given the benefit of standardizing technology, in order to represent a child's work accurately.

If a school accepts externally developed criterion-referenced tests, it must also accept the curriculum on which the tests are based. The test results read something like: "knows how to do long division; does not know how to divide fractions." The tests themselves are "secure," so that parents and students cannot interpret the results for themselves.

The problems of this kind of standardized testing are still being explored.

The National Assessment of Educational Progress is a federally funded national program. It gives standardized tests to samples of students and others aged 9, 13, 17, and 26 to 35 on a number of educational programs.

The tests are recycled every three years to determine trends in educational progress, according to the following schedule.

Cycle I		Cycle II	
1969-70	Citizenship, Science, Writing	1975-76	Reading, Literature
1970-71	Reading, Literature	1976-77	Music, Social Studies
1971-72	Social Studies, Music	1977-78	Math, Science
1972-73	Math, Science	1978-79	Writing, Career and Occupational Development
1973-74	Writing, Career and Occupational Development	1980-81	Reading, Literature
1974-75	Art, Citizenship		

The results will not help school districts directly as other standardized tests may do. They are reported for the nation as a whole, and for four regions — never for states or school districts.

Rather, the trends shown in repeated cycles are intended to help educators in assessing educational needs in a very general way. For example, knowledge of science — which is the only area that has been tested three times — has been found to have declined seriously in these last ten years.

Those who take the tests respond, naturally, from their whole experience — not from their school experience alone. The ages of the children tested do not correspond exactly to grade levels. Thus the test results could conceivably be of value to other than educators. Publishers of various types of materials, for instance, might find the scores very helpful.

The same tests are available to states for assessment by sampling of their own students. This is a possible alternative to other sorts of standardized testing in that (1) individual students cannot be identified; (2) the time burden of taking the same set of tests can be divided among several students of the same age; and (3) data are available from the national assessment for comparisons.

HOW ACCURATE ARE TEST SCORES LIKELY TO BE?

Norm-referenced achievement tests purport to tell us how any group of students compares with other groups of students in the same grade. School systems use the norm of each group as the basis of comparison. How accurate is this comparison likely to be?

A number of factors affect individual test scores. For example:

- The assumptions underlying many test items may be

foreign to a child's thinking and experience. Test-makers do look at widely used textbooks in constructing tests. Nevertheless, students taking a test in Iowa, Appalachia, and inner-city Boston almost inevitably have had different instruction and widely different experiences in life.

For instance, a group of inner-city children were baffled when they were asked whether a lark was a cheese, a bird, an automobile, or a dog. They all knew it was a widely advertised cigarette. Other children might have known it was both a bird and the name of an automobile. Which should they choose?

•The time allowed to complete the test is limited. The score a child achieves depends largely on how much of the test is completed within a given time, since all incomplete answers are scored as incorrect. Thus a child's speed in doing the test could be as important as giving correct answers.

Teachers in most schools coach students in how to take the tests. "Work fast," they say. "Fill the squares neatly — computers won't score a right answer poorly marked. And most of all, don't think too hard!"

•The wording of many test questions is ambiguous. Because of this, imaginative students tend to do badly. Place yourself in the position of the child choosing answers in the following typical examples of questions in standardized tests:

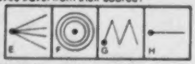
1. An orange seed grows into:
   a. an orange tree
   b. an orange
   c. another seed
   d. an orange blossom

(Science Research Associates, Achievement Series, Blue Green Series, 1968)

2. In warm weather, water changes into water vapor and rises into the air. Later this water vapor changes into rain or snow and falls to the earth. This process is called:
   a. evaporation
   b. condensation
   c. precipitation
   d. the water cycle

(Stanford Achievement Test, Primary Level, 1972)

3. Which one of these drawings most nearly shows the way sound waves travel from their source?



(Metropolitan Achievement Test, Intermediate Level, 1970)

In each case a thoughtful student could find reasons to check any one of the answers; but a standardized, computer-scored test cannot take into account the reasoning of a student. Nor does it pretend to examine how much a student knows about any given subject. All the test does is tell whether the student gives what the computer has been programmed to accept as the answer to a given question.

Perhaps in recognition of the fact that *no* question which asks for any degree of analytical thinking can have just one answer, large numbers of standardized tests ask for the *best* answer rather than for *the* answer.

For example, a reading test in the current California Achievement Tests (CAT) describes the success of two students, Margo and Bert, in planting a garden in a vacant lot later preempted by the city for a parking lot. The students surveyed the neighborhood people for their views on what had happened.

Several questions with multiple-choice answers include the following: "What will the students *probably* do with the results of the survey?" and "Which of the following words *best* describes Margo?"[3]

Despite the latitude suggested by the questions — which could as well have been essay questions — the computer still considers only one answer correct. Students sometimes complain that the multiple-choice answers offered in such questions omit equally plausible or even better answers. It is not clear to them why a "test" should not require correct answers.

- Another major problem for children is that many items intended to test a particular skill also require special knowledge. For instance, the following is an excerpt from a third-grade reading test:

A *cinder* is a piece of __________

rock   sand   firebrick

burned coal[4]

Is the child who does not know the answer therefore unable to read?

- The tests all too frequently exhibit insensitivity to the feelings of non-Caucasian children — an insensitivity that affects the self-esteem of the children and thus their learning progress.

The following two questions reflect this insensitivity, and also use language generally unfamiliar to children.

When a dove begins to associate with crows, its feathers remain __________ but its heart grows black.

a. black
b. white
c. dirty
d. spread
e. good

(Lorge-Thorndike Multi-Level Edition, Level 3, Verbal Battery, p. 3)

How the __________ roses flush up in the cheeks!

a. white
b. pretty
c. small
d. yellow
e. red

The "correct" answer is "red" — but only if the cheek in question is white. *(Same Lorge-Thorndike test.)*

- Children think differently from adults. A Stanford-Binet subtest for ten-year-olds asks two questions concerning "why children should not be too noisy in school" and "why most people would rather have an automobile than a bicycle."[5] Quite often the answers the ten-year-old would pick would be different from what an adult might expect.
- Some tests reflect other "cultural bias." The Wechsler Intelligence Scale for Children (WISC) General Comprehension subtest asks, "Why are criminals locked up?" "Why is it better to pay bills by check than by cash?" "Why is it generally better to give money to an organized charity than it is to a street beggar?"

These questions may well puzzle some children from very poor families who have seen relatives or friends jailed for reasons they did not understand. They may also puzzle children who have never heard of a bank account, or have never seen evidence of organized charity.[6]

Children who grow up in more or less the same culture as the test-makers do far better on standardized tests than do children from a different culture.

Numerous studies show a strong correlation between parents' education and income levels and student scores. The president of the Educational Testing Service was quoted in the *Boston Sunday Globe* of November 12, 1978, as calling this "a bias of opportunity."

- Other factors that may affect how well a child does on a standardized test — other than knowledge of the concepts tested — are the attitude of the examiner to different children in a class, an attitude which frequently expresses an expectation that one child will do well and another poorly; whether or not the child had breakfast; the child's reaction to time pressure; the child's view of the importance of the test; and the child's tolerance of ambiguous or trivial questions.

WHAT CAN STANDARDIZED TESTS TELL US?

What can standardized tests actually tell us about what children have learned?

- The objectives of standardized testing are generally limited to content of facts, definitions and their applications, and events in various areas of language, mathematics, science, and social studies. The test scores tell us something of the child's past experience or instruction in relation to the curriculum concepts on which the test was constructed, and something of the child's recall of that experience.
- The scores give some measure of the skills required to deal with the test. They tell us something of the child's neatness, facility for guessing, and ability to work reasonably well under a time limit.
- The tests can tell us little about the child's ability to learn, to analyze, or to reason — and nothing of the child's judgment, originality, imagination, or creativity. Nor do the tests tell us how much a child knows about any given subject. Rather, tests simply skim the surface of learning.
- Standardized tests are of no use in diagnosing individual student needs. Test areas are too limited. Further, the test questions are "secure," and returned to the test-maker. The student or parent can never check back over the questions when the scores are returned.

In some instances schools can ask analyses of the questions to be returned with the scores, but that is not the same thing as having the questions themselves. Moreover, the scores are returned weeks and even months after the test was taken.

- There are wide margins for error in the individual scores themselves. For instance, a student would be likely to give at

least some different answers on the same test if it were given on different days. This affects what is called the statistical "reliability" of the test.

Test-makers try to reduce the likelihood of different answers — that is, to increase the test's reliability — by giving similar tests to the same students at different times and by comparing their answers for a "reliability coefficient." They remove from the test items on which students are likely to change their minds in favor of a group of items for which the "coefficient" is acceptable.

A reliability coefficient of .95 is said to be excellent, .90 acceptable, and .80 "not so good."

Some of the most widely used tests are based on a reliability coefficient of .90, adjusted for an average score of 500 points with an allowable "standard deviation" of 100 points. This deviation allows a range of test "error" that is more than 100 points wide.

That is, if a child's score is 550, it is "90 percent sure" that his or her true score is somewhere between 500 and 600.

In percentiles, however, as scores are recorded, this means that a score of 500 may be reported at the 50th percentile, and a score of 600 at the 84th percentile — a major difference to a child, the child's teacher, and the child's parent.

At the same time, the scores of the other 10 percent of the students will lie outside this range altogether.[7]

For such reasons many test-makers themselves urgently warn school systems not to use standardized-test scores for evaluating individual students, stressing that the tests are meant only to measure group progress. On the other hand, others give instructions for interpreting tests for individuals.

A task force of the National Council of Teachers of English warns parents that using national tests "to judge an individual student or to compare the students of a single class is like using the scales in a truck-weighing station to measure whether everyone in the family gets the same amount of ice cream."[8]

Almost universally, however, the schools have badly abused the tests by using them to categorize students.

The scores are made a key part of student records. Parents who have checked their students' record files sometimes find standardized test scores and nothing else.

Test scores are used to characterize the abilities of students for parents; to make class assignments; to group or "track" students according to supposed ability; to serve as a basis for guidance in selecting courses and setting career goals; and to stand as a prime factor in awarding scholarships.

Following the example of the schools, parents often give test scores far too much importance in judging their children. They give emphasis to "fact" learning, which has relatively little staying power, and tend to ignore development of values and understandings, which contribute to the growth of the child.

These abuses of test scores put students at risk of being permanently mislabeled and misdirected, while millions of parents and students remain unaware that the scores are so badly misused.

THE DILEMMAS OF STANDARDIZED TESTING

Education faces a number of dilemmas in using standardized testing — dilemmas that deeply concern test-makers, and necessarily involve parents and teachers.

- Because the tests offer such an easy way of comparing large numbers of students, school systems generally have come to look upon standardized tests as a way in which they may give an accounting of their stewardship to the public. Yet, if the tests are unfair to students, are they fair to the schools?
- School systems ordinarily compare their norms with the norms of other systems. But a norm is an average — not a goal. If a national norm becomes a goal, the school is aiming for mediocrity.
- Standardized test scores across a school system can give curriculum planners for that system some suggestions of curriculum needs, by comparing local instruction against the national norms. The scores, for example, may suggest a review of subject matter or a change in teaching methods and materials.

On the other hand, the norms on which the tests are based relate to national samples of students. As we have seen, the assumptions underlying the tests may not fit the local students in motivation, experience, cultural background, or other important ways. Any use of the tests must take such differences into account.

- A series of tests can lay out a sequence of skills by grade levels, which is helpful in building curriculum devoted to the development of reading or mathematical skills.

At the same time, standardized instruction does not always provide for the many individual differences in the ways children learn — differences that are still the subject of intensive research.

- Standardized-test scores can locate potential problem areas among large groups of students. Other tests are needed, however, to determine the nature of the problem. Still other kinds of assessment are needed to help individual students.

- Because of the near universality of their use, standardized tests are credited with raising the level of education across the country — particularly for minorities — by defining a set of common learnings.

This assumption could be more easily accepted if the tests were not kept secret by the terms of the contract with the schools that use them. If the learning level of students is to be assessed according to their test scores, then the desirability of the questions must also be assessed — and no such general study is now possible.

- Observers say that increases in reading test scores in some urban school systems appear to be related to extensive practice with workbooks similar to the achievement tests. It is well known that coaching for standardized tests is prevalent in many — if not in most — school systems.

This poses a question about the reliability of test scores as indicators of academic achievement, as opposed to an increased competency in taking tests.

Yet the credibility of standardized test scores is such that they are commonly used as reliable indicators of student achievement in assessment of new programs, school reorganization, and a host of other instances.

For example, an Associated Press dispatch in *The New York Times* of December 24, 1978, reports the success of ten years of integration in the schools of Berkeley, California, as assessed by significant improvements in standardized achievement test scores among both black and white students.

- Standardized testing has long been closely related to the age-grading system by which most schools are organized. Horace Mann used it, among others, in seeking ways to improve education. In the nineteenth century and later, also, there have been times when promotion was a political issue. Standardized testing gave an "outside authority" to the teacher in determining whether or not a child should be promoted to the next grade.[9]

The concept of "learning disability" appeared at this time. Students with whom a teacher did not succeed could be transferred to a special class.

Although the grading system has persisted as a convenient way of managing large numbers of children, most authorities agree that children differ in the amount of time they need to achieve a given level of learning.

Children commonly find themselves at different grade levels in different subjects. A fifth grader may easily be at the fourth-grade level in math and at the eighth-grade level in reading. Teachers frequently complain that graded textbooks do not fit their classes, and seldom use the same text for all students.

The emphasis on standardized testing adds to the problem that Murray Levine has so aptly described:

"The imposition of a constant time for all children to learn a given amount has been a psychological disaster for untold millions of children and their parents and for teachers who had to make agonizing decisions."[10]

- As an item of achievement is mastered by more and more students, it becomes less useful as a test item in measuring performance. As a result, test-makers turn to different ways of writing the same curriculum item. They complicate the wording, or they phrase the question in the negative. For example:

Which of the following is not a real factor in the growth of slum areas?

( ) disease
( ) unemployment
( ) crime
( ) climate
( ) don't know

(Metropolitan Achievement Test, 1959)

Teachers object to this practice. They say that a negatively phrased question often becomes a test of logic, rather than a test of the instructional objective.[11]

- The limitations that scoring-by-computer places on testing are the subject of widespread debate. On the one hand, the multiple-choice format is said to be "objective," whereas scoring of essays or responses to open-ended questions is said to be more "subjective."

On the other hand, putting every question into a multiple-choice format restricts the content of the curriculum covered, as well as the degree of thought required of the student.

English teachers, for instance, note weaknesses in the quality of standardized tests to measure such areas as appreciation of literature; sensitive use of language; organization of ideas; critical reading and listening; and expression of values, among others.[12]

Social studies teachers want curriculum to "provide young people with resources for making decisions affecting their public and private lives as citizens"[13] — resources that include skills and attitudes that are not measured by achievement tests.

Teachers are looking for tests that will meet these needs, tests that will reflect cultural diversity and measure actual teaching objectives. At the same time, test-makers are searching for ways to assess the reasoning and creative abilities of students. This is slowly bringing about a long-needed collaboration between test-makers and educators concerned with curriculum.

These are questions that deeply concern PTA people as well. Their investment in their schools is too great for them not to be involved in the ways tests are used.

- Standardized tests directly affect curriculum development, textbooks, teacher education, admission to higher education, and the dispensation of scholarship funds.
- Beyond the outlay of school funds for the tests themselves, the tests also cost heavily in staff time; in learning and teaching time; and in heavy emotional pressures on students, teachers, principals, and parents.
- Moreover, the exaggerated importance given standardized tests raises grave questions about the directions in which they are taking our children, their parents, and their schools.
- Most students will meet the standards we set. If the most important tests in students' lives are hinged upon the average, the level of our schools tends to mediocrity. Are we asking of schools what we really want of education?
- There are many kinds of tests, and tests can be used in different ways. If, because of the tests we use, the advancement of students in school depends always on measuring themselves against their age-mates, students tend to place a low value on themselves and their own differences.
- On the other hand, what really matters in a child's education cannot be measured by any test existing today. Stephen Musseau put well the basic need:

"What are the steps for building self esteem in children? Help your child see that he is beautiful and unique just as he is! Guide him away from believing that he must stack up to others! Success is a journey, not a destination!"

IQ TESTS ARE STANDARDIZED TESTS

Intelligence or IQ testing started with Alfred Binet's experiments in France at the beginning of the century. His purpose was to find a basis for exempting children who did poorly in school from the compulsory attendance laws, or to substitute special classes.

Binet himself was skeptical of his work. He said that the idea that "the intelligence of an individual is a fixed quantity" is a "brutal pessimism."[14]

Others, however, seized upon the notion that children's intelligence could be measured. In the United States, intelligence tests were first used to test the poor for "feeble-mindedness," to assign them "a suitable place" and for "curtailing reproduction."[15]

In these decades large numbers of psychologists and others talked about "an aristocracy of talent" to be developed through sorting people according to ability. Sterilization laws were passed in several states.

Later a massive testing program was used to weed out "undesirable aliens" among immigrants and to set immigration quotas for various countries of origin. These quotas were based on the United States Census of 1890. Because the quotas had been filled at the time of need, they were directly responsible for the deaths of hundreds of thousands of victims whom Adolf Hitler was willing to deport.

During World War I some 125,000 draftees were given IQ tests in order to determine who could most easily learn certain tasks. One finding noted that "foreigners who had lived in the country twenty years or more before being tested were every bit as intelligent as native Americans. Those who had lived in the country less than five years were essentially feeble-minded."[16]

These days are long past; yet the same ideas still turn up in newspapers, in school conversations, and even in published research. It was the tests in World War I that first offered massive "evidence" that nonwhites scored lower on IQ tests than did whites. In the last decade, books asserting the same theories about the "inferior intelligence" of non-Caucasians claimed wide attention—theories also based on standardized tests developed with Caucasian children.[17]

Walter Lippmann wrote in the 1920s:

"If the impression takes root that these tests really measure intelligence, that they constitute a sort of last judgment on the child's capacity, that they reveal 'scientifically' his predestined ability, then it would be a thousand times better if all the intelligence testers and all their questionnaires were sunk without warning in the Sargasso Sea."

The Oxford English Dictionary defines "intelligence" as "the faculty of understanding." The American Heritage Dictionary defines it as "the capacity to acquire and apply knowledge." We speak of "native" intelligence, meaning an inborn quality or characteristic.

"IQ" stands for "intelligence quotient"; this is defined technically as the "ratio of tested mental age to chronological age, multiplied by 100."

Mental age, then, is not what a youngster was born with but something acquired over time through experience or instruction. IQ tests measure learning that has been acquired.

The following are typical items used in IQ tests:

1. How tall is the average American man?
Any answer from 5'7" to 5'11" is acceptable, but none even a half inch more or less.
(WISC-R Manual, p. 68)

2. Something you see in your sleep is a __________
( ) dream ( ) fairy
( ) wish ( ) dread
(Iowa Tests of Basic Skills, Primary Battery, Level 7, Form 5. Vocabulary Subtest)

3. A handbeater : an electric mixer :
broom : vacuum cleaner
flashlight : light bulb
sink : dishwasher
wrench : vise
(Cooperative School and College Ability Tests, Series II, Form 3B, p. 3)

4. Where there is __________ let me sow love.
( ) suspicion ( ) friendliness
( ) hatred ( ) love
( ) hope
(Cognitive Abilities Test, Form 1, Levels A-H, Verbal Battery, p. 11)

5. dollar peso mark lira ______
( ) change ( ) franc
( ) foreign ( ) purchase
( ) bank
(Lorge-Thorndike, Level 4, Verbal Battery, p. 6)

A half century after Lippmann's alarmed warning, then, available evidence does not show how intelligence tests differ from achievement tests — in which children are supposed to make progress. One observer collected ten questions, half from group ability (IQ) tests and half from achievement tests. Can you tell the difference?

1 We see (children, plants, stars, houses, trees) only at night.

2. Bill bought two pads of paper at 25 cents each and four pens at 20 cents each. How much did he spend? ($1.30, 45¢, $1.04, none of these)

3. *Sob* means (prejudice, solemn, sigh, joy, kind).

4. When a new kind of machine is created it is called (an adoption, an invention, a fabrication, a novelty, a discovery).

5. Which term is missing in this series? 3, 5, 7, 11, 13? (8, 9, 10, 14, 15)

6. Which term comes next in this series? 54, 45, 36,? (31, 63, 25, 27)

7. Oxygen is a (compound, gas, solid, carbide).

8. To *prove* is to (agree, verify, see, mean).

9. Mary bought a comic book for 10 cents, some gum for 5 cents, and a candy bar for 5 cents. How many cents did she spend in all? (15¢, 20¢, 25¢, 50¢, none of these)

10. The earth's crust is its (surface, energy, heat, poles).[10]

The statistical methods used by test-makers to determine "mental ages" for intelligence tests are similar to those used to determine norms on the grade-level achievement tests.

This means that untold numbers of youngsters have been labeled with irrelevant test scores. They were measured, not for native intelligence, but for the information they happened to have acquired in their lives. But the labels ordinarily stand for the duration of their years in school, and often affect their lives.

Only in recent years have protests mounted against the use of group IQ tests — especially when it became clear that such factors as better nutrition, native-language teaching, and more perceptive instruction raised IQ scores that had been supposed to be fixed.

Group IQ tests are used less often today. Individual tests of various sorts are often part of more intensive student evaluations for diagnosis of learning problems.

Some test-makers are still convinced there must be a way to measure intelligence defined as cognitive aptitude — a way that would be different from academic achievement tests; but they admit it has not yet been discovered.

EVERY CHILD IN A BELL-SHAPED CURVE

Both IQ tests and achievement tests are shaped so that scores will fall into an arbitrary mold called the "bell-shaped curve." This is the theoretic basis of the standardized test scoring described earlier.

Test-makers — and many educators — are convinced that the abilities of most students, and therefore their test scores, run close to the average; those above and below the average thin out on either side. When the scores are plotted on a graph, therefore, the curve should follow a bell shape, with the scores of highest frequency represented by the top of the curve, and the sides of the curve rising and falling smoothly.

It is true that when a group of *measurable chance* events are recorded on a graph, the likelihood is that they will follow the bell-shaped curve.

For instance, the measured blood pressure of a sample of general population was found to take a bell shape, indicating that most people are close to an average or norm.

On the other hand, the sizes of families are not chance events and do not take a bell curve. Neither do the sizes of family incomes. Both of these, plotted on a graph, rise sharply at the beginning of the graph (the low end) and fall off gradually at the other end.[11]

In all three of these cases, measures can be exact. Blood

pressures are recorded. Family sizes and family incomes are verifiable numbers.

It has been said that the incidence of native intelligence ought to be similar to the incidence of blood pressure. Possibly so: but until some reliable measurement of intelligence has been identified, there is no basis for assuming that children's intelligence will fall into a bell curve.

Nevertheless, test-makers construct achievement and IQ tests as though the test scores were equivalent to measures of intelligence and on the assumption that such measures would take a bell curve.

After tests have been tried out with samples of students, items are chosen (and scoring is determined) according to the way the scores fit the curve — called the "normal distribution." Items that do not bring about the curve are excluded, or the scoring is changed.

Parents and teachers should use the knowledge of this background to keep their children's scores in perspective. Neither IQ nor achievement scores should ever be taken alone as describing a youngster's abilities, or allowed to override careful day-to-day perceptions. Basing all expectations of a child's abilities on so imperfect a measure borders on child abuse.

WHAT ABOUT THE SATs?

When the decline in SAT (Scholastic Aptitude Tests) scores broke into print two years ago, it seemed to focus for many people a pervasive uneasiness about public education. At the same time, however, it strengthened an already existing concern about testing.

SATs are standardized tests, but they are different in several ways from the standardized IQ and achievement tests we have been discussing.

SATs have only one purpose: to predict the likelihood of success in college, particularly in the first year. SATs have been used since the 1920s to help college admission people sort out applicants whose grades were earned at widely different high schools.

More than a million seniors take the SATs each year — about half of the number who actually go on to college, and about one quarter of the 18-year-old age group.

The SATs are "performance" tests. That is, the scores reported are the scores students actually earn by performance on the tests, not by reference to a national sample. SATs have two parts: verbal and mathematical. Scores on each part run from a base of 200 to a possible 800.

The verbal portion of the test is designed to assess reading skills and understanding of word relationships. It covers four areas: antonyms, analogies, sentence completion, and reading comprehension. The material for this is drawn from social, political, scientific, artistic, philosophical,and literary writings. For example:[30]

(1) Choose the word or phrase most nearly opposite in meaning. (antonyms)

Babble: (A) irrigation (B) pollution (C) meaningful speech (D) useful object (E) helpful person

(2) Choose the pair best expressing a similar relationship. (analogies)

Choir : Singers:
(A) victory : soldiers
(B) class : teachers
(C) crowd : protesters
(D) challenge : duelists
(E) orchestra : musicians

(3) Choose the word or set of words that best fits the meaning. (sentence completion)

Intricately carved and beautifully proportioned, the priceless sculpture was __________ the work of a __________
(A) understandably . . . dilettante
(B) indelibly . . . forger
(C) demonstrably . . . bungler
(D) unmistakably . . . master
(E) paradoxically . . . perfectionist

(4) The student is given a passage of some 400 words which may be summarized as follows: (reading comprehension)

The behavioral sciences are developing rapidly in understanding, prediction, and control of human behavior. At this time the public is ignorant or disbelieving of its possibilities, just as in 1900 the public rejected the possibilities of aeronautics. At a future time the public will accept and use the findings of the behavioral sciences.

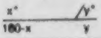
Among questions based on this passage is the following: The author suggests that the next change in the public's attitude toward behavioral science will lead the public to:
(A) ignore the findings
(B) increase the use of the findings
(C) disbelieve the findings
(D) use these findings against each other
(E) lose interest in the findings

The mathematical portion of the test requires as background mathematics typically taught in grades one through nine; it depends less on formal knowledge than on reasoning. It measures students' problem-solving ability in three areas — arithmetical reasoning, elementary algebra, and geometry. For example:

---

1. The number which is one hundred less than one million is:
   (A) 990,900 (B) 900,900 (C) 999,990 (D) 999,900
   (E) 1,000,900

In each of the following two questions, choose:
- A if the quantity in column A is greater
- B if the quantity in column B is greater
- C if the two quantities are equal
- D if the relationship cannot be determined from the information given:

	Column A	Column B
2.	$5x + 3y = 22$	
	x	y
3.		
	$180-x$	y

How serious was the drop in SAT scores to American education?

In 1952 the average verbal test score was 476, and the average mathematics test score 494. In 1963 — the high point — the averages were: verbal — 478; math — 502. In 1977 the averages were: verbal — 429; math — 470. In the last year or two the decline appears to have leveled off.

The 49-point drop in verbal scores is 10.25 percent; the 32-point drop in math scores is 6.37 percent. In terms of percentages these declines do not seem remarkable in themselves. But educators were concerned not only because of the steadiness in the decline — however slight — but also because the decline increased in the last three years. Since the educational performance of students has a tremendous influence on our country's future, educators, parents, and the youngsters themselves wanted to know if the decline reflected upon the students, the schools, or our society.

The College Entrance Examination Board appointed a "blue ribbon panel" (often called the Wirtz Committee from the name of its chairman, Willard Wirtz, chairman of the National Manpower Institute) to study the implications of the declining scores. Its report is called *On Further Examination*. It is not a long report and deserves PTA reading.

A summary cannot do justice to its thoughtful writing, but for our present purposes we can extract these findings:

- The panel was convinced by its studies that the tests themselves are still consistent in predicting ability to do college work; that the decline in scores did not result from changes in the test or in the methods of scoring it.

At the same time, the panel urged a broader look at the whole picture of testing: Are we testing in the best possible way whatever ought to be tested?

- High school grades are still the best single predictors of college performance; but grades combined with SAT scores give still more accurate predictions. Even so, these predictions are far from perfect. The best median coefficient for SAT scores and high school grades combined is .56 — which means that just over 30 percent of the students perform as predicted.
- Part of the decline in SAT scores seems traceable to the opening up of opportunities to attend college. Fewer dropouts, less discrimination, and more choice in colleges played a role in the decline. Another factor was the less than adequate national effort to prepare less advantaged students for educational opportunity as it became available.
- The panel felt, however, that no more than one quarter of the decline since 1970 could be attributed to the larger number of students taking the tests. The other three quarters of the later decline apparently resulted from the impact of more "pervasive" forces — in the schools, in several major societal changes, and in student motivation.
- The panel's "firmest conclusion" was (1) that less thoughtful and critical reading is now demanded in high school, and (2) that careful writing has apparently gone out of style. The report suggests that "increasing reliance . . . on tests requiring only the putting of X's in boxes contributes to juvenile writing delinquency."

Restoration of critical reading and careful writing is of central importance.

- The panel "conjectured" that math scores declined less because math is less dependent on the verbal skills; also, that math is essentially school-based learning, while verbal skills are more influenced by experience in the home.
- Absenteeism in the schools is another probable cause. "To the extent school is a place youngsters flee because they find themselves diminished or bored or both, the need is plainly for internal reform. Yet without the right alliance with home and community the school's effectiveness is limited."
- The panel probed at length such conditions in schools as grade inflation, social promotion, reduction in homework, and the "watering down" of textbooks to lower grade levels. It concluded that there has indeed been a lowering of educational standards and that this is a major factor in the decline.

But the answer, they said, is not that education must be more rigid — rather that education must become still more diversified to meet the needs of all students, but without being watered down:

"The situation requires the collaboration of teachers, students, parents, and the broader community in the establishment of standards that can be considered higher only as they recognize youth's essential diversity."

- Home learning is of critical importance, the report says. Both changes in educational practice and changes in family life-styles have probably contributed to a strain on the teacher-parent relationship. That strain in itself may have had more impact on students than anything either partner may have done separately.
- Too little is known about the impact on achievement of the "hanging" state of childhood and adolescent development.
- The report quotes E. B. White to the effect that television is the test of the modern world. It is probably one cause of the SAT score decline. At the same time, television has the potential of making great contributions to the process of learning.
- The greater declines between the years of 1972 and 1975 suggest strongly that disruption in national life in the immediately preceding years affected the preparation of these students adversely.
- The apparent decline in student motivation suggests that "during the past 10 years the curve of the SAT scores has followed very closely the curve of the entire nation's spirits and self-esteem and sense of purpose."

The postscript is that in 1978 math scores went down two points and verbal scores held firm at last year's 429.

What happened?

High schools contend that more attention to writing produced higher scores.

At the same time, coaching for SATs by outside agencies has increased to the point of provoking a federal investigation. This in turn increased the controversy between the test-makers' advice to students that cramming does not help and the documented pattern that students who take cramming courses achieve substantially higher scores."

The end of the decline — at least temporarily — suggests the common human response to the squeaky wheel. It does not answer the "values questions" that have been raised about how we can best use testing to help students learn — or the differences between high scores and education.

THE COMPETENCY MOVEMENT

Public reaction built slowly. Over the past three years rising dissatisfaction with schools, school graduates, school taxes — or all three — triggered a movement called minimum competency testing. It spread through practically all states with unprecedented speed.

- Business people complained that graduating students came to them functionally illiterate and arithmetically incompetent. High school graduates found themselves unable to find and apply for jobs; some filed suit against their schools for poor instruction.
- Colleges instituted remedial classes for students who had acceptable SAT scores but had never learned to write.
- Minority-group citizens widely felt their youngsters were being shortchanged. Minority-group youth suffered even higher unemployment rates than youth of other groups.
- Inflation and rising taxes focused taxpayer resentment on schools. Declining test scores in SATs, in the National Assessment of Educational Progress, and in standardized achievement tests gave substance to such complaints. People said that high school diplomas had become meaningless.
- The media gave more than ample attention to all of these.

Legislators responded to constituent pressures. It was a good political issue. State educational agencies rushed to secure their defenses. In the majority of states, either legislatures or state boards of education have already initiated or are planning to initiate some sort of statewide testing of every student.

For instance, the California legislation uses language common to bills of other states in requiring that, beginning with the class of 1980, "pupils attending public schools in California acquire the knowledge, skills, and confidence required to function effectively in contemporary society."

With such promises, the "instant popularity" of competency testing even reached the point that bills were introduced in Congress to provide for national competency testing. To date, support is small, but the possibility of some national program is not to be dismissed. (The National PTA Board of Directors has expressed its opposition to any federal testing program.)

BASIC SKILLS: THE NEW EDUCATIONAL GOAL

In the previous decade a number of states had been working quietly in various patterns of state-local cooperation to revise curriculum through setting goals and translating objectives into local classrooms in an effort to "fit" local student populations.

The competency movement took the opposite course. It said, essentially, that schools were too slow or were not sufficiently earnest in their endeavor — certainly that the public was unconvinced that schools were actually improving. If the students were learning basic skills, let the schools prove it by testing every child.

What has happened in the last three years is reminiscent of the man who jumped on his horse and rode off in all directions.

- The pioneer states simply made competency testing a requirement for high school graduation. The hasty action confused the schools, but they tried to comply.
- In 1972 Oregon required local districts to define their own requirements in each of 13 categories: reading, writing, speaking, listening, analyzing, and computing, plus seven "survival skills," such as being an informed citizen, being an informed consumer, and maintaining health.

One district adopted 203 separate competencies in these categories; another, 143. They included driver training, using a telephone, writing checks, and making a peanut butter sandwich. The resulting nightmare of paperwork — and parent indignation — cut the number of competencies drastically.

The tests remained relatively easy. The 1978 graduates — the first group to meet the requirements — turned in varying failure rates of 1 and 2 percent.

At the other extreme, Florida made statewide tests a graduation requirement for all students. The Educational Accountability Act, passed in 1976, specified that tests in "functional literacy" and mathematics be given to all eleventh graders without delay. This allowed a short year's grace for "remediation" before diplomas were awarded or withheld.

The tests given a few months later required the customary passing grade of 70. Across the state 36 percent of the eleventh graders failed the math tests and 8 percent failed in reading and writing.

The tests were criticized for cultural bias. The school system was charged with inadequate instruction of minority-group students, because 79 percent of black students failed in mathematics and 26 percent in communications.

The tests were most criticized because both schools and students felt they had been given insufficient warning that receiving a diploma would depend on passing this test.

Reaction ranged from a substantial increase in student dropouts to widespread praise of the state administration for moving toward higher school standards. The following year the state department of education could point to a better record. In 1978 an estimated 73 percent of eleventh graders, statewide, passed the math tests, against 64 percent in 1977; and 97 percent passed the communications tests, against 92 percent in 1977.

Nevertheless, in October 1978 the law was challenged in Federal District Court as unconstitutional. The complaints were discrimination; that the skills examined were not the skills taught; and that the test was not a valid or reliable instrument — meaning that it had not been properly pilot-tested among a cross-section sample of students.

- Policy-makers across the country observed the experiences of these and other pioneer states. A host of questions began to emerge, and among them three conflicting and troubling problems:

1. how to devise programs that would satisfy political demands and at the same time address fairly the very real weaknesses in schools;
2. how to impose universal standards of doubtful levels upon extremely diverse school populations without destroying contemporary achievements in individualized education;
3. how to foresee, if possible, and take into some account the long-range implications of these actions for retaining excellence as an educational goal and for the related needs of a democratic society.

WHAT COMPETENCIES ARE MINIMUM?

Nobody really agrees on what every child should know. It is generally agreed that virtually all youngsters need skills in communications and math. Beyond that there are significant differences of opinion.

In the last two decades the balance of employment has been reversed to the point that less than 5 percent of present jobs are unskilled. Some time ago school administrators accepted as a goal the idea that every student must graduate from high school "with a salable skill." The standard of the competency movement appears to be less than that.

- Basic skills in communications and math are the major part of the elementary curriculum. Mastery is expected by the end of the eighth grade. Some feel that schools should emphasize the mastery of academic skills and that students should be able to apply mastered skills to life situations as they arise. Available competency tests indicate, however, that some state and local school systems are moving in a different direction. They are testing for "life" skills. For example:

Florida tested academic skills cast in "real-world" terms. Some of its specified communication skills were:

---

a. Determine the main idea inferred from a selection.
b. Find the who, what, where, which, and how information in a selection.
c. Distinguish between facts and opinions.
d. Identify an unstated opinion.
e. Use an index to identify the location of information requiring the use of cross references.
f. Use highway and city maps.
g. Accurately complete forms used to apply for a driver's license, employment, entrance to a school or training program, insurance, and credit.

Mathematics tests were based on questions such as the following:

---

a. Determine the elapsed time between two events, stated in seconds, minutes, hours, days, weeks, months, or years.

b. Determine equivalent amounts of up to $100, using coins and paper currency.

c. Determine the solution to real-world problems involving purchases and a rate of discount given in fraction or percent form.

d. Solve a problem involving capacity, using units given in a table (milligrams, grams, kilograms, metric tons, ounces, pounds, tons) for conversion within the system.[22]

The following comes from a Vermont graduation test:

If someone buys a $5.38 item from you and gives you a $10 bill, what change will you return?

Critics point out that some youngsters are able in math but not in reading skills; nevertheless, most "problems" test students for reading first, and mathematics second. For instance:

A company has a 2.5 meter high wire fence around its equipment storage area. The area measures 30 meters by 50 meters. The company plans to add extra security to the area by stringing 3 strands of barbed wire along the top of the fence. How many meters of barbed wire does the company need for this job? 160, 240, 480, or 960 meters?[23]

Some competency programs include as "basic skills" knowledge of consumerism, the democratic process, practical science, health maintenance, social responsibility, personal and career development. For example:

Last year you joined a record club. You fulfilled your obligation to buy the specified number of records. You called the record club and asked to be removed from the mailing list. The club continues to send records anyway. Which of the following actions would be appropriate for you to take, and which would not?

a. Keep the records and refuse to pay any bills for them.

b. Send a registered letter asking to be removed from the mailing list.

c. Pay for the records that came after your phone call, and call again requesting that the company stop sending records.

d. Refuse to accept the records and have the post office return the records to the company.

(National Assessment of Educational Progress, 1977)

Another example:

Which of the following would you expect to find in a democratic society?

1. Joe Smith gives voters $5 each to vote for him.
2. Citizens legally picket and protest a court decision.
3. A group of people go to the city council to ask for an investigation of the mayor.
4. Congress overrides a Presidential veto.
5. A citizen is arrested for breaking a law that is not written down.

(Westside High School, Omaha, Nebraska)[24]

These items suggest possible questions. Should schools generally be expected to teach methods of dealing with life problems? Can they cover the basic laws and regulations involved in the day-to-day life of an adult — or a representative part of them? Should this take precedence over other curriculum?

Second, is it reasonable to expect that whatever skills are tested in a minimum competency program are attainable by all students during school years? Are these and similar tests "minimum" learning? Or are they desirable learnings? How much of such content could reasonably be left to application of basic skills and common sense in later life?

Third, available competency tests of "life skills" apply to today's world. What of the future? Credit practices change. Laws change. Schools have difficult choices to make in responding to what they hear as popular demand, and in judging how to spend precious time.

Generally missing in competency tests is the usual academic material of secondary schools—history, literature, science—the background of lifelong learning. Generally missing, also, are the social skills that enable people to function in a democracy; the personal skills that determine the quality of life for an individual; and even minimal knowledge of the rest of the world on a planet that grows smaller every day.

Does this foreshadow gradual change in school curricula? Schools generally teach what students are expected to need for tests that both the school and the student regard as important — an understandable protection for both.

WHAT IS FAIR?

When will students have had adequate instruction in the content on which they are tested?

Due process principles established by the courts have slowed down the competency movement in one respect: Adoption of competency tests as graduation requirements seems to have stopped with 13 states. The present trend, moreover, is to give tests required for high school graduation to ninth graders first.

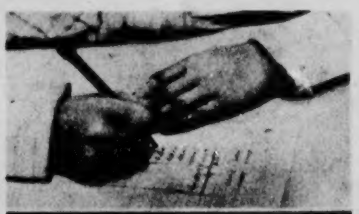
In many states it is provided that tests be repeated until each student has had every possible chance to be successful. Some states require remedial programs.

- Some states are recognizing that high school is late to correct problems in basic skills and are beginning to focus on the early elementary grades.

A dozen or so states require testing at least four times throughout the 12 grades. A few have required use of the tests for grade promotion; others intend them only for local school guidance. In many states local school systems have devised their own competency programs. There are no consistent patterns.

- It seems to be just as difficult to determine a fair cutoff requirement for achievement levels as it is to choose the competencies to be tested.

Henry M. Brickell has described the experience of a group of high school teachers who devised tests in math and English for the end of tenth grade. The tests appeared to him to be of about fifth grade difficulty. A passing score of 60 was required. About 25 percent of the tenth graders flunked the English section; about 50 percent flunked the math.



Afterwards he asked, "Suppose remediation doesn't work. How many diplomas can you withhold?" "Maybe 3 percent," they said. "Then 97 percent have to pass the competency tests. What level of work can you *guarantee* that 97 percent of all graduates can do?"

"First-grade work — maybe second grade, if you mean a *guaranteed* minimum."[25]

This does not mean that some tenth graders are doing all second-grade work. It means that some students vary widely in their abilities in different areas. It means that some schools will have a difficult time in finding a "least common denominator" for all students.

- There was a time, before the present age-grading system became established, when teachers apparently expected children to progress at different rates.

What schools are hearing in the politics of the competency movement, on the other hand, is that the community expects all students to reach certain standards at certain points in their educational programs.

This raises very difficult questions for school administrators: What is the expectation of the community for student achievement? On the other hand, what percentage of failure will the community accept? If the level of failure is high, is the community prepared to pay for the necessary remedial programs? If not, will the community accept the likely increase in dropouts — students who want to avoid the admission of failure — with the well-known social costs?

For example, the New York State Regents withdrew one set of minimum competency tests when the triviality of many items drew public ridicule and substituted a more difficult set to be effective in 1981. Yet even the old tests are expected to deny some 7,000 New York City seniors their diplomas this year.

Large numbers of these students are Hispanic. Community organizations charge that the students did not receive transitional help.[26]

The city of Baltimore undertook to reexamine its entire program when two-thirds of its students failed the first round of tests.

Some school districts in the Southwest, with large numbers of students learning English as a second language, reported a failure rate of 50 percent.

Some school districts have arrived at acceptable compromises. For example, Denver has maintained for more than 18 years a competency program that places competency requirements for graduation at about a sixth-grade level. It also maintains a strong remedial program. Its failure rate of graduation has run consistently at 2 to 4 percent.

- A high percentage of success in the competency tests, however, does not of itself mean that the program is fair to students.

In any school one minimum level may be too hard for a slow student, laughable for a quick student, and unmotivating for both.

On the one hand, the old Gresham's Law of economics — that cheaper currency drives out hard money — seems to be warning that the lower standards of minimum competency testing could drive out higher academic standards. Some students who have passed basic math tests have actually been counseled against taking further math courses in high school.[27] Should not passing the competency tests have argued competency for the next level of math courses?

On the other hand, ninth-grade testing has caused schools to place academically weak students in lower "tracks" in preparation for the tests.

Those who fail, in the words of one Phoenix school administrator, "are so loaded with remedial courses they have no chance to get into humanities or other courses that might keep them in school."[28]

Even if such students finally succeed in passing competency tests, their high school years have been spent in the bottom tier of the "two-tiered" education that hopelessly divides so many children for life — with little or no attention to their own talents or education for growth.

Along with the dropouts, and with the youngsters who receive a "certificate of completion" instead of a full diploma — because they passed courses but failed competency tests — these students will find it extremely difficult to establish their competence for postsecondary education or for desirable employment.

It is ironic that many low-income groups, such as Title I councils, see the competency movement as a hope for securing needed attention for their children. They have not seen that their children will bear the brunt of competency testing as now generally constituted; that the effort to make the schools accountable may actually push their own children out of school.

WHAT DOES COMPETENCY TESTING COST?

One Virginia school superintendent reported a cost of half a million dollars to establish competency tests for a school population of some 150,000 children. A preliminary report from Oregon listed direct costs ranging from $26,500 to $173,200 per district to devise plans and establish programs.

Costs for keeping records and reporting results must follow the establishment of a testing program. These have been estimated at as much as $25,000 per year per district.

Costs of remedial programs could be heavy, especially where there are large numbers of minority-group students. Washington state has estimated remedial costs of $90 to $100 million. New Jersey estimated a cost of $70 million for the current year. Michigan is spending some $28 million.

Developing new tests periodically averages about $100 per student. Multiply this by, say, a minimum of 20 million students a year. If tests are to be made available to students and their parents after they are given and scored, new tests will have to be constructed more often.

State department budgets will include costs for either developing and administering test programs or providing assistance to local districts. These costs are variously estimated from $65,000 to $495,000 annually.[29]

Nobody really knows how much competency testing will cost in school funds. And indirect costs are also possible.

Many teachers are concerned that basic skill remediation in the secondary grades may displace curricula in humanities and the sciences. And it is the lack of the higher order skills, they say — the real basics of critical thinking — that is the actual source of employer discontent and college dissatisfaction. Sooner or later, these are costs that will have to be reckoned with.

WHAT FORECAST FOR EDUCATION?

For the most part, the competency movement was thrust upon the schools as an attempt to repair our neglect of large numbers of children — particularly minority-group children, but also many others who have been slipping through schools without learning needed skills.

Even in this short period, the competency program has provided plenty of evidence of the needs of children. Meeting these needs is another matter. By itself, competency testing can do nothing — whatever sanctions are imposed. Testing has underscored weaknesses in basic-skills learning that were generally known in schools but more or less hidden and ignored, for various reasons.

The danger now is that hasty action will leave these students virtually untouched, for the influences that insist upon mass testing and are inclined to make mass judgments of schools are also generally unaware of the processes by which children learn.

Only instruction can make a difference; and only dedication to reaching every child in the early grades can be substantially successful.

The best possible outcome of the competency movement can be the broad national discussion it has provoked of the goals and means of education.

If the competency movement leads to a serious look at poor teaching and other problems in learning; if it leads to more direct instruction in basic skills in early grades; if it leads to more communication among teachers, more "cross-over" in teaching the same basic skills in all courses; if it leads to more consideration of what is important to teach, and involves the total school community in these considerations — then the competency movement could be a "watershed" for American schools.

On the other hand, competency testing is an instrument with its own dangers to schools.

Goals of mediocrity can scarcely restore credibility to education. A competency program implies — however unjustly — a sort of guarantee of performance in basic skills on the part of those who have passed its tests. The direct invitation to "tracking" students can lead only to a division of our young adults by education that will be no solution to employment problems and will only deepen our social problems.

The political popularity of competency testing could lead from state-initiated testing to state-controlled — or possibly a national — curriculum. The outcome depends, perhaps, on both the efforts of schools and the commitment of communities.

Education is never going to be a neatly ordered, deliverable package. No "mass" remedy ever has been successful in dealing with school problems, because they are essentially problems of individual differences. Competency testing is no exception.

WHAT TO LOOK FOR IN YOUR SCHOOL'S TESTING PROGRAM

- Whatever the assessment program is in your school, parents should learn about it. They should make themselves aware of the purpose of different kinds of tests in a child's educational program. They should be able to see the questions or other directives after a test, in order to monitor the student's progress.
- Parents should be wary of allowing others or themselves to exert pressure on students in relation to tests. Tests should not be too numerous, or loom too large in a child's school life, or parents and teachers will find themselves "pulling up the carrots to see if they are growing."
- No one test is a fair basis of assessing a student's progress. Individual scores are at most a minor indicator of a child's progress — never a permanent indicator of ability — and should never be used as the basis for any administrative decision.

The school probably uses standardized achievement tests of some sort. These scores should be placed alongside a child's classwork; scores on teacher-devised and other criterion-referenced tests; the teacher's estimate of the child's abilities; and the parent's knowledge of his or her talents and interests.

It should be understood that a child's abilities, activities, and attitudes cannot be measured.

- Parents and teachers need to look at their children's education as a whole — a single continuum from kindergarten through twelfth grade — with a systematic evaluation of student progress. The only true estimate of a child's progress can be based on change over a period of time.

NOT ALL CHILDREN MATURE AT THE SAME RATE

The "assembly line" grading system ignores the realities of children's uneven growth patterns. Knowing that children differ in their talents and strengths, parents and teachers can insist that instruction be related to children's growth patterns; that testing be based on instruction; and that continuous efforts be made to improve teaching and learning.

It is not reasonable to expect strong competencies in all skills from all children. Very often, moreover, development of individual strengths also brings solutions to weaknesses. For instance, children often learn to read more readily when their talents in one of the arts or their interests in sports are recognized.

- Before choosing among the many available alternatives for assessment, schools should make a sustained, in-depth examination of what teachers need to know about the process and content of their students' learning. This study should involve both students and parents.

Giving parents and students a voice in curriculum development and assessment is an effective way for a school system to fulfill its responsibility to account to the community for educational performance. Such a process would earn schools new public support — as well as help to attain a better quality of learning for every child.

- It is most important that emphasis on basic skills start in early grades. Helping teachers to monitor individual programs from the beginning could go farther in strengthening these skills than a uniform testing program. If a testing program is adopted, however, it also should begin in the primary grades as part of instruction.

- In the case of secondary students who have failed to acquire reasonable skills, parents and teachers should insist that remedial measures be supportive, not demeaning; that they be appropriate to the age level of the student and conducive to the development of self-esteem; and that they leave room for other instruction on the secondary grade level.
- Teachers should be encouraged to use a variety of approaches. Flexibility in ways of measuring skills — the use of other methods and not just pencil-and-paper testing — helps the teacher help the student. "Test" is too restrictive a word.

Some schools use a packet of student papers, journals, interviews, talks to the class, artwork, problem-solving tasks in community projects, consumer tasks, use of math in construction or cookery, and a host of other examples of practical application of basic skills.

Some states, such as Vermont, have directed local districts to have teachers "check off" skill achievement for each child as it occurs in day-to-day class work. It is well known that any child — like any of us — will often need to learn a skill more than once; if the check-off task is not too burdensome, it makes the point that achievement has occurred.

- Where the objective of a testing program is success for every student — rather than ranking or sorting — the manner of test writing is important. There is every reason why tests should be straightforward, without tricks or negatives.

Writing skills should be judged from writing samples, with the important skills of editing included in the task. That is, students should be given an opportunity to rewrite a piece. Editing skills are a large part of writing skills, and more often ignored than not. Teachers do well in evaluating a student's writing by looking first for content, credibility, clarity, and organization — and then for punctuation and other mechanics.

"Holistic" scoring is a skill that many teachers are learning. Pilot tests in the Massachusetts competency program have shown that teams of teachers can score large numbers of student papers with strong agreement in their evaluations and with a surprisingly small expenditure of time.

Oral communication should be judged from interviews and debate as well as prepared talks, according to such criteria as focusing on the topic and speaking clearly and logically. Listening and viewing skills are being given new importance in a world growing dependent on oral and visual communication, and the need for performance tests in these areas is obvious.

Besides these alternatives, teachers can turn to various criterion-referenced tests — whether their own or externally developed — to check on mastery of particular subject matter. Unambiguous multiple-choice questions can fairly be used to test specific instruction, such as grammar or computation.

- More open-ended assessments, however, give a better idea of how a student learns. In science and social studies, the use of "What do you know?" questions — as well as "Do you know what I know?" — removes the crutch of multiple-choice answers, because such questions afford the child an opportunity to show comprehension of a sequence of ideas and reasoning abilities.

Giving the child free choice in the library for independent reading can suggest his reading level — as well as developing curiosity and an adventurous mind.

- Teachers can develop their students' own abilities to evaluate progress by asking them to correct each others' papers, and then by following up with a general discussion.

Teachers can also lead students to raise questions about literature or history — and then to evaluate those questions as (1) interesting, (2) average, or (3) trivial.

Improving student powers of evaluation through such practices also improves learning. Students should be continuously involved in documenting and evaluating their own progress.

- Alternatives to paper-and-pencil tests recognize that individuals differ in the styles in which they learn and communicate — that some are more confident in writing, some in speech, some in artistic expression. The famous composer Bruckner once expressed his difficulty in responding to honors paid to him by saying, "If I had an organ I could tell you how I feel."

If the objective is competency in learning "to function effectively in contemporary society," it seems that students need not attain the same level in every skill at the same time. Attaining skills enough to feel the "confidence" envisioned by the California legislature — along with an eagerness for learning — may be the best foundation.

In the end we have to recognize that there is no way to test what really matters to a child's learning. Every child is a unique individual. Children share some abilities, have some talents of their own, and differ in the way they learn.

As one educator put it, a student is not a fire to be lighted, or a vessel to be filled, but a being whose growth is to be nourished.

18

NOTES

1. Sherwood Davidson Kohn, "The Numbers Game: How the Testing Industry Operates," *The Myth of Measurability*, ed. Paul H. Houts (New York: Hart Publishing Company, 1977), pp. 162-68.
2. Ibid.
3. California Achievement Tests, Book 16C, pp. R6-R7.
4. *Reading Tests: Do They Help or Hurt Your Child?* North Dakota Study Group on Evaluation (Grand Forks, North Dakota, n.d.).
5. Lewis M. Terman and Maude A. Merrill, *Stanford-Binet Intelligence Scale: Manual for the Third Revision* (Boston: Houghton Mifflin Company, 1960), p. 96.
6. Amado M. Padilla and Blas M. Garza, "A Case of Cultural Myopia," *Myth of Measurability*, p. 131.
7. Mitchell Lazarus, "Coming to Terms with Testing," Ibid., pp. 185-89.
8. *Common Sense and Testing in English*, National Council of Teachers of English (Urbana, Illinois, 1975), p. 32.
9. Murray Levine, "The Academic Achievement Test: Its Historical Context and Social Function," *American Psychologist*, March 1976, p. 232.
10. Ibid., p. 231.
11. *Common Sense*, p. 21.
12. Ibid., p. 16.
13. Janet Hanley Whitla and Dean K. Whitla, "Social Studies Tests," *Myth of Measurability*, p. 276.
14. A. Binet, *Les Ideas Modernes sur Les Enfants* (Paris: Flammarion, 1913), pp. 140-41.
15. Lewis M. Terman, *The Measurement of Intelligence* (Boston: Houghton Mifflin Company, 1916), pp. 6-7.
16. C. C. Brigham, *A Study of American Intelligence* (Princeton University Press, 1923), p. 100.
17. Mitchell Lazarus, "A Second Look at Jencks's 'Inequality,'" *Myth of Measurability*, pp. 324-30.
18. Judah L. Schwartz, "The Illogic of IQ Tests," Ibid., pp. 91-92.
19. Philip Morison, "The Bell-Shaped Pitfall," Ibid., pp. 83-86.
20. Material of this section is drawn largely from *On Further Examination. Report of the Advisory Panel of the Scholastic Aptitude Test Score Decline*, College Entrance Examination Board, New York, 1977.
21. Mitchell C. Lynch, *The Wall Street Journal*, December 3, 1978.
22. Thomas H. Fisher, "Florida's Approach to Competency Testing," *Phi Delta Kappan*, May 1978, p. 600.
23. *Miami Herald*, January 6, 1978.
24. Jim Findley, "Westside's Minimum Competency Graduation Requirements," *Phi Delta Kappan*, May 1978, pp. 614-18.
25. Henry M. Brickell, "Seven Key Notes on Minimum Competency Testing," Ibid., pp. 589-92.
26. *The New York Times*, January 12, 1979.
27. Barry Mitzman, "The Problems of Minimalcy," *The Math Learning Center Report*, Math Learning Center (Salem, Oregon, 1978).
28. *The Competency Challenge: What Schools Are Doing*, An Education U.S.A. Special Report (National School Public Relations Association, Arlington, Virginia, 1978), p. 73.
29. Barry D. Anderson and Philip Lesser, "The Costs of Legislated Minimum Competency Requirements," *Phi Delta Kappan*, May 1978, pp. 606-8.

A PARENTS' GUIDE THROUGH THE TESTING MAZE

Parents, teachers, and students will learn more from tests given at their school if they understand that:

there are different kinds of tests with different purposes;

all tests are limited in what they measure and what they can tell about the student; and

misuses of tests can be harmful to students.

How does a "standardized" test differ from tests that teachers give to their students periodically to check on instruction in their own classrooms?

A standardized test can be any test that large numbers of students take under similar conditions. The *norm-referenced achievement* test, however, is the most widely used standardized test.

These norm-referenced achievement tests do not evaluate so much what students may be studying in school at the time but rather evaluate general "constructs" — reading comprehension, number concepts, or computation.

The scoring of norm-referenced achievement tests does not represent what a student *should* accomplish, but rather a range of what students on a given grade level ordinarily *do* accomplish.

Why should parents be concerned about the ways in which standardized tests are used in their schools?

Standardized tests directly affect curriculum development, textbooks, teacher education, admission to higher education, and the dispensation of scholarship funds.

Standardized tests are costly — in outlay of school funds for the tests themselves; in staff time; in learning and teaching time; and in the heavy emotional pressures they may put on students, teachers, principals, and parents.

Standardized tests are being given exaggerated importance. Consequently, grave questions are raised concerning the directions in which they are taking our children, parents, and our schools.

If the most important tests in students' lives are hinged upon the average, as standardized tests are, then the level of our schools will tend to mediocrity.

If, because of the tests we use, the advancement of students in school is determined by comparing them with their classmates of the same age, will we not destroy their self-esteem and their appreciation of individual differences?

A number of states have mandated competency testing as a requirement for high school graduation. Out of the experiences of these states have emerged a host of questions and at least three conflicting and troubling problems:

How to devise programs that would satisfy political demands and at the same time address fairly the very real weaknesses in schools;

How to impose universal standards of doubtful levels upon extremely diverse school populations without destroying contemporary achievements in individualized education; and

How to foresee, if possible, and take into some account the long-range implications of these actions for the related needs of a democratic society and for retaining excellence as an educational goal.

Below is a checklist of questions parents might ask when they examine their own school's testing program.

What kinds of tests are being used and for what purposes?

Ask to see the tests. Do they have a geographical, cultural, sexual, or racial bias? Are the questions clear? Do the questions require special knowledge? Are the questions straightforward, and not designed to trick students rather than test them honestly about what they have been taught?

How much pressure is being put on students — by parents or by members of the school faculty — to do well on tests?

Have the test publishers warned that "grade levels" are not absolute standards but rough indicators of averages?

Has your school cautioned against using test scores to label students or label teachers, and has it warned the media that test scores are unreliable information on the quality of a school?

Are standardized test scores being used alone, or, as they should be, along with a child's classwork; scores on teacher-devised and other criterion-referenced tests; the teacher's estimate of abilities; and the parent's knowledge of talents and interests? One test can offer at most only a minor indication of a child's progress.

Is the education of your child viewed as part of a continuum, kindergarten through twelfth grade?

Is a systematic evaluation of your child's progress conducted, based on this continuum?

Is instruction related to the individual growth patterns of each child?

And is the testing based on that instruction?

Are continuous efforts being made to improve teaching and learning at your school as a result of testing?

Did your school conduct an in-depth examination of what teachers need to know about the process and content of their students' curriculum before choosing the tests to be given?

Are parents and students given an opportunity to have a voice in curriculum development and assessment?

Does emphasis on the basics begin in the early grades? Do the teachers monitor individual progress in these grades? Does testing begin in these early grades as a part of the instruction?

Is remedial instruction supportive rather than demeaning? Is it appropriate to a student's age level and conducive to the development of self-esteem? Does it leave time for other instruction and activities?

Are teachers encouraged to use a variety of testing methods?

Are appropriate performance tests given in the areas of writing and oral communication?

Are some of the tests given open-ended?

Do the tests allow for differences among students in styles of learning and communicating?

READING LIST

Ahmann, Stanley. *How Much Are Our Young People Learning? The Story of the National Assessment.* Fastback 68. Bloomington, Indiana: Phi Delta Kappa Educational Foundation, 1976.

Brickell, Henry M. *Let's Talk About Minimum Competency Testing.* Denver: Education Commission of the States, 1978.

Common Sense and Testing in English. Urbana, Illinois: National Council of Teachers of English, 1975.

Compact, Winter 1978. Denver: Education Commission of the States.

The Competency Challenge: What Schools Are Doing. An Education U.S.A. Special Report. Arlington, Virginia: National School Public Relations Association, 1978.

Evaluation as Feedback and Guide. Yearbook, Association for Supervision and Curriculum. Washington, D.C.: Association for Supervision and Curriculum, 1967.

Glass, Gene V. *Standards and Criteria* (Monograph). University of Colorado, n.d.

Haney, Walt, and Madaus, George, "Making Sense of the Competency Movement," *Harvard Educational Review*, November 1978.

Houts, Paul L., ed. *The Myth of Measurability.* New York: Hart Publishing Company, 1977.

Miller, Barabara Soloth, ed. *Minimum Competency Testing: A Report of Four Regional Conferences* jointly sponsored by the Education Commission of the States and the National Institute of Education with the cooperation of the Carnegie Corporation of New York. St. Louis, Missouri: CEMREL, Inc., January 1978.

Minimum Competence, Parts One and Two, *Options in Education*, June 5 and 12, 1978. Washington, D.C.: National Public Radio and the Institute for Educational Leadership of the George Washington University.

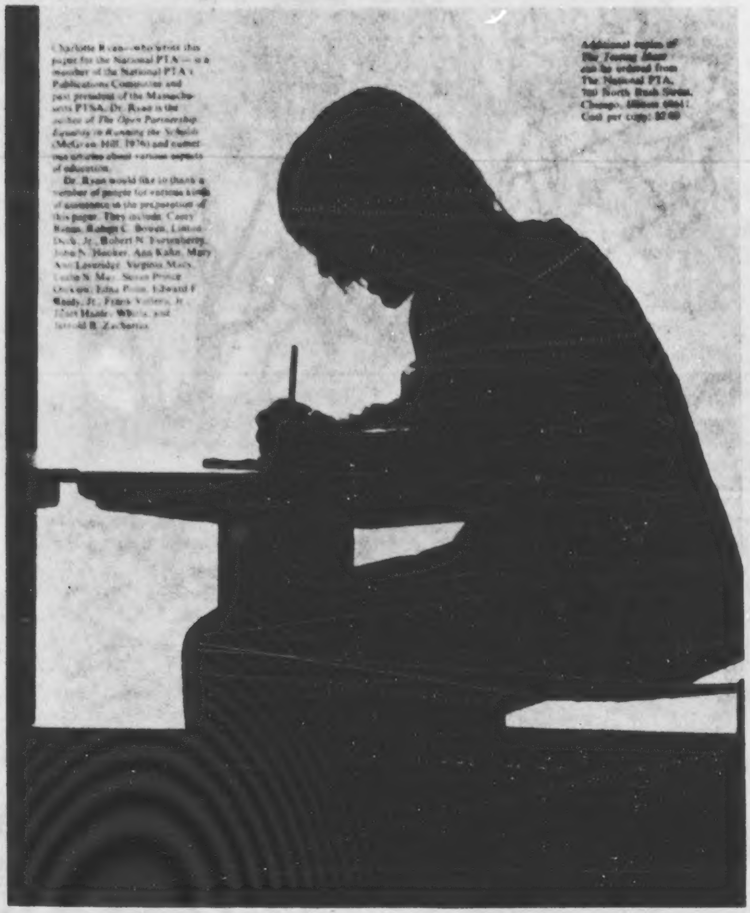
National Elementary Principal, July/August 1975, October 1977, June 1978, January 1979. National Association of Elementary School Principals, Arlington, Virginia.

On Further Examination. Report of the Advisory Panel of the Scholastic Aptitude Test Score Decline. New York: College Entrance Examination Board, 1977.

Phi Delta Kappan, Special Issue on Minimum Competency Testing, edited by Chris Pipho, May 1978, Bloomington, Indiana.

Zasloff, Barbara. *A Bibliography on Bias in Intelligence and Achievement Testing of Children and Youth.* North Dakota Study Group on Evaluation, January 1976.

124-1279—2.00



Charlotte Ryan—who wrote this paper for the National PTA — is a member of the National PTA's Publications Committee and past president of the Massachusetts PTSA. Dr. Ryan is the author of *The Open Partnership: Equality in Running the Schools* (McGraw-Hill, 1976) and numerous articles about various aspects of education.

Dr. Ryan would like to thank a number of people for various kinds of assistance in the preparation of this paper. They include: Casey Banas, Ralph C. Brown, Linton Deck, Jr., Robert N. Fortenberry, John N. Hocker, Ann Kahn, Mary Ann Leveridge, Virginia Mack, Leslie N. May, Susan Prince Osborn, Edna Prine, Edward F. Ready, Jr., Frank Vollero, Jr., [illegible] Wheeler, and Jerrold R. Zacharias.

Additional copies of *The Testing Maze* can be ordered from The National PTA, 700 North Rush Street, Chicago, Illinois 60611. Cost per copy: $2.00

Mrs. Sparling. As parents and educators we support H.R. 4949 as a means for requiring the disclosure of information which will address the eight concerns just raised and we feel that the following are hallmarks of the legislation.

(*a*) That the bill provides for the standard error of measurement to be shared not only with teachers but with test takers and their parents. Many of them will, for the first time, be aware of the broad range of precision evident in these scores, even though life and career decisions are being made on the basis of a few points.

(*b*) That the legislation includes protections to the right of privacy of family and test takers in insuring that research will be limited to data which is not personally identifiable, unless approved by the test taker.

(*c*) That need to release to test takers their actual test questions and answers within a reasonable period following the tests. Also important is a procedure for review of a challenged test score within a reasonable time frame.

We are hopeful that the section which will require a description of special services to accommodate handicapped students will allow schools to deal more effectively with the needs of such students for differentiated testing.

We are finding locally, for example, that many students with writing difficulties can be appropriately tested by oral tests. To try to force written tests on such youngsters is cruel and unproductive.

In closing, I would like to emphasize three points:

(1) the charge has been made that these proposed changes will dramatically increase the cost of testing. While data are available to challenge that statement, this would be a poor excuse for maintaining a system of testing which is falling under serious challenge. Schools and families now spend over a quarter billion dollars a year on testing, and the appropriate question might well be: Are we already wasting $250 million?

We cannot evaluate the changes that may be necessary unless we can review what the true costs of the present system cover.

(2) testing is a large private for profit industry which has a tremendous impact on a public phenomenon, education. The public has a right to scrutinize this industry and to hold it accountable for its product.

(3) the National PTA does not seek an end to testing. But we are seriously concerned that the present degree of the art of assessment is far less precise than we have been led to believe.

We feel the questions about our present testing methods will not be put to rest until parents and teachers are able to make decisions based on much more data than is now available to us. That is why this legislation is so important.

We seek the information that will give all of us dependable data to help determine the course of our public schools and the appropriate direction for our children.

The National PTA stands ready to participate in any reasonable effort to reach that goal.

Thank you.

Chairman Perkins. Thank you very much.

We will now hear from Dr. Rever.

Go ahead, Dr. Rever.

STATEMENT OF PHILIP REVER, DIRECTOR, WASHINGTON OFFICE, AMERICAN COLLEGE TESTING PROGRAM, ACCOMPANIED BY RICHARD FERGUSON, VICE PRESIDENT OF ACT'S RESEARCH AND DEVELOPMENT DIVISION

Dr. **Rever**. Mr. Chairman, I am Dr. Philip Rever, director of the Washington, D.C. office of the American college testing program. With me is Dr. Richard Ferguson, vice president of ACT's Research and Development Division. We are pleased to share with you and your subcommittee our statement concerning H.R. 3564 and H.R. 4949.

ACT is a not-for-profit corporation which develops and publishes numerous assessment programs administered to students in behalf of the Nation's secondary and postsecondary schools.

One such program, the ACT Assessment, is used nationwide to facilitate students' transition from secondary to postsecondary education programs. Accordingly, ACT would be subject to the provisions of either of the two bills currently under consideration by the subcommittee.

We do not believe that these bills are in the best interests of the Nation's students. Neither do we believe that they serve to remediate any pressing problems directly attributable to the uses of tests in this country. Consequently, we respectfully urge that the bills not be enacted.

We have prepared quite an extensive statement for the committee which we have submitted to the committee, and I would like to summarize our reasons for the opposition for your two bills.

Our comments are directed specifically at H.R. 4949, although it is our intention that they be extended to the corresponding provisions of 3564 as appropriate.

The testimony contained herein deals with both the general but highly significant, educational and political issues that have motivated and enveloped discussions of the two bills, and with the specific shortcomings of the bills.

I will review for you this morning in brief form what our general thoughts are about both bills rather than a section by section analysis.

We believe that the proposed legislation is ill-advised due to many factors, some political, some educational, and some simply reflecting a misunderstanding of the truly significant educational issues of the day.

We regret that so many resources and so much energy are being invested in efforts to place unneeded controls on the tools that have served and continue to serve students and education so well.

We lament this occurrence, not out of any sense of self-interest, but rather in recognition of the futility of focusing disportionate attention on the tools that repeatedly remind us of one of our society's most repugnant shortcomings, that we have not yet succeeded in assuring equal eduational opportunity for all of our peoples.

To believe that legislation requiring the dissemination of all manner of information about tests will in any way remediate basic societal ills, some of which have been scores of years in the making, is to engage in the most fanciful and wishful thinking.

If, on the other hand, a goal of the legislation is to diminish the value of tests, then the bill is well suited to this purpose.

We do not believe this to be the goal inasmuch as the realization of such a goal would ultimately result in the sweeping of one of our Nation's most urgent problems under the rug. It is for this reason that we encourage a more intensive and extensive study of the potential long-term effects of the provisions of the bill.

We are further troubled by the lack of awareness on the part of so many as to the potential negative effects of H.R. 4949 for students, educational institutions, and ultimately the public which depends on a well-educated citizenry for its future well being.

The prevailing attitude in some quarters that enactment of the proposed legislation need not be preceded by a thorough study of the potential impact of its provisions, because these provisions are at worst discomforting to test agencies and test sponsors, belies the reality that many provisions of the bill have potential consequences that are counterproductive to the attainment of the stated goal of the bill.

Let me review for you briefly what we believe the major significant negative consequences of H.R. 4949 are.

ACT identifies positively with the stated purpose of H.R. 4949. We do not believe, however, that the provisions of the bill contribute in significant ways to the attainment of that purpose. To the contrary, we view many of its provisions as potentially counterproductive. Specific illustrations of the shortcomings of the bill in this respect are provided next.

Essentially we have divided these effects into two categories: (1) political and legal, and (2) educational and social.

With regard to the political and legal effects, enactment of H.R. 4949 will preempt responsibilities reserved by the U.S. Constitution for States. Although two States have enacted legislation similar to H.R. 4949, only the law enacted in the State of New York is comparable to the proposed legislation.

The bill enacted in California is very dissimilar from H.R. 4949. Of particular note is that two States, Maryland and Colorado, have considered similar legislation but rejected it. Moreover, some States conduct intrastate admission testing programs which they develop without the assistance of any testing sponsor or agency.

Thus, enactment of H.R. 4949 will impose the will of the U.S. Congress over the will of the elected representatives of at least two States and extend Federal jurisdiction to testing programs solely within the province of State education agencies without compelling reasons.

It is noteworthy that, to our knowledge, no legal challenge has ever been brought involving a test sponsor's violation of the Constitutional guarantee of due process and equal opportunity. Specifically, we are unaware of any court of law having decided in favor of a challenge to the validity and appropriateness of tests used in a college admissions process. This fact speaks strongly to the lack of a compelling reason for Congress to enact H.R. 4949.

H.R. 4949 contains provisions that violate the spirit and intent of existing Federal law. Sections 4(a)(2) and 5(c) would appear to rescind proprietary rights previously granted by the Congress to test sponsors under copyright law.

Sections 4(a)(1)B and section 7 are contrary to an historic intent of the Federal Government, maintaining competition by protecting trade secrets. By enacting H.R. 4949, the Congress will thus have reversed protections granted under previous Federal legislation such as the Freedom of Information Act and the Securities and Exchange Act.

Withdrawal of copyright and trade secret protection from a specific group of agencies is discriminatory and contrary to the large public interest.

With regard to educational and social effects, because there are marked differences among admissions tests insofar as their purposes, theoretical foundations, content, uses, and the magnitude of those uses, H.R. 4949 will have differential impacts on the various admissions tests. That is, some programs will be more seriously affected than others.

For example, programs with low volumes, and even those with large volumes, will find it difficult to offer the complete services presently offered to students and educational institutions. In particular, special testing provisions may be severely hampered.

Special testing arrangements for persons with physical handicaps or with religious preferences that forego testing on the regular test dates will be affected. Some testing agencies also provide institutions with the facility to test students on-campus at times of the institutions's choosing. Such testing would have to be restricted or discontinued altogether.

The release of secure test items as required by the bill will have numerous negative effects of an educational character. Primary among these is the expected impact of the scope and content of the tests on high school curriculums.

The inevitable result of releasing test items will be to encourage and/or facilitate many schools and teachers to strive to increase student performance on the tests, not always in a pedagogically sound manner. To the extent that this occurs, tests will be inappropriately leading the curriculums.

A related effect may well be to encourage the proliferation of coaching schools, many of which will fail to engage in sound educational practices. Since it is not established that the efforts of such schools produce results of equal benefit to all students, any such outcome of the bill would be anathema to the public interest.

We believe that educators and educational institutions are more likely to be successful in serving society if they make informed decisions about the educational process, whether regarding individual students or programs affecting those students.

Likewise, students are more likely to achieve their fullest potential if they are well informed about their capabilities in relation to the requirements of various educational and vocational programs they may be considering, and about ways of readying themselves for those programs.

Some of the tests affected by the bill are designed to accommodate such informed decisionmaking. Indeed, the value of the tests to students and educators in this context far exceeds any possible value of them in the selection decisions attributed to them.

Given these uses of tests, it is indeed lamentable that they should be subjected to provisions which can only diminish their perceived value to students, educators and the public.

The provision of H.R. 4949 that would require the release of test items will likely have such an effect.

The release of test items will almost certainly subject the items and the tests to highly inappropriate analyses. The probable lack of recognition that different types of tests are constructed to serve diverse purposes will likely result in invalid interpretations and critiques of the items by individuals unschooled in testing methodology and uninformed about the intended uses of the tests.

Such occurrences will result in a diminishing of the confidence placed in tests for all uses. When this occurs, the loss to our educational system will be significant; the loss to the individuals who are helped by the test data, either directly or indirectly, will be even greater.

In summary, we have provided in our statement comments on H.R. 3564 and H.R. 4949 as part of what we believe to be our responsibility to American education. It is our conviction that the proponents of this legislation, while well-intentioned, have concocted a remedy whose ultimate consequences will likely overshadow the illness supposedly being addressed.

We are concerned that the very important reach of these bills within our educational system, both in terms of pedagogy and in terms of the proper role of government, has been given short shrift.

It is crucial that subsequent study of the bills by the Congress emphasize a thoughtful assessment of the need for this legislation balanced against the impact it could have. The potential negative consequences, to some extent motivated by lack of understanding of the nature and purpose of the tests themselves, surely deserves more serious attention than these hearings will provide. To that end, we strongly urge thorough subsequent study of the testimonies provided.

We are dealing with the potential of literally millions of persons. It is undeniable that tests play an important role in this context. That role is appropriate in the educational setting and has been found sound in the occupational setting by the U.S. Supreme Court.

This legislation jeopardizes the contribution of tests to our educational system and ultimately poses grave problems for the autonomy of our institutions of higher learning.

The burden of proof regarding the need for legislation on standardized testing cannot be met by political sloganism or populist concepts put forth without benefit of data that verify the nature and extent of the harm proponents of these bills assert applies.

Until such data are provided, the proposed legislation should not be enacted.

Mr. Chairman, we are here to answer your questions, and I have provided the committee with a catalog of our publications. Should the subcommittee so decide, we will add them.

[The remainder of Dr. Rever's prepared statement follows:]

STATEMENT OF PHILIP REVER, DIRECTOR, WASHINGTON OFFICE, AMERICAN COLLEGE TESTING PROGRAM

SHORTCOMINGS OF H.R. 4949

We have discussed above general reasons why we do not believe H.R. 4949 is an appropriate means of addressing either its stated or implied objectives. This judgment resulted, in part, from a careful study of the Bill's provisions, and the reluctant conclusion that they frequently reflect an incomplete understanding of well established testing methodologies and programs. This observation is best illustrated by specific commentary on each provision of the Bill.

Section 2.(a)(5)

Contrary to the Bill's declaration that no "proprietary interests" of the testing sponsors will be abridged by the Bill, we are not convinced. (See our comments on Sections 4(a)(1)B, 4(a)(2), and 5(c) for amplification.)

Section 2.(a)(6)

This provision appears to establish interstate commerce as the motive of Congress in governing a process within the academic domain. Education has traditionally been the province of the states, within specific constitutional limits. Congress should consider carefully the close relationship between how an institution of higher learning determines its role in society and the means by which it chooses its students. Despite statements to the contrary by the sponsors and proponents of the Bill, the spectre of government control of or involvement in academic institutions' rights to self-determination lies well within the realm of this proposed legislation. Though governments frequently encroach upon this ill-defined territory, it is clear that a line must be drawn defining the extent to which the federal government, in particular, should become involved in academic processes.

Legislation dealing with matters similar to those in H.R. 4949 is in effect in California and New York. Numerous other states have considered legislation and found it to be undesirable. In this connection, the federal government is well advised to consider what compelling reasons exist for imposing its particular views on these states. Furthermore, the Bill extends federal jurisdiction to intrastate test programs that are developed and administered by state education agencies. We believe the legislators and citizens in these states should have the right to determine for themselves the appropriateness of such legislation as is proposed.

Given that the need for and the appropriateness of federal action are highly questionable, then such action seems to be an abrogation of rights left to the states by the U.S. Constitution.

Section 2.(b)

We endorse these purposes. We have contributed to their fulfillment for many years. We call attention to the large gap between these purposes and the Bill's actual provisions in subsequent passages of this testimony. We emphasize at this point, however, the following:

Section 2.(b)(2)

The inclusion of the word "appropriate" to qualify this provision is highly significant and deserves strict adherence. For example, if strictly applied, we believe that the Bill's sponsors will be unable to supply evidence and reasonable rationale for the release of secure test items. In view of the harmful impact that such an action will likely have on the educational process, the action would be highly inappropriate.

Section 2.(b)(3)

H.R. 4949 will not substantially protect the public interest through the requirements it sets forth for testing sponsors. The standards and practices of the testing sponsors are presently higher than any ascertainable outcome of the Bill's provisions. Further, the assumption that legislation will result in "greater accuracy, validity, and reliability . . ." of the tests involved is highly speculative. In fact, the proposed legislation will almost certainly have the opposite effect for reasons given later in this testimony.

Section 2.(b)(4)

We strongly support what we view to be the intent of this provision. We believe it is improperly framed for two reasons, however. First, it appears to assume that 'other' elements of the "multiple criteria" postulated in this provision are viewed by Congress as equally or more valid and reliable than standardized tests. The facts belie this assumption and are available in considerable detail should there be interest in exploring them.

Second, the provision seems to suggest that many criteria are necessarily better than one -- without regard for the validity of the criteria involved. This is most inappropriate. Indeed, we challenge those who appeal for inclusion of such variables as creativity, compassion, sensitivity, motivation, etc., to produce evidence of the validity and reliability of available measures of those attributes.

Section 3.(a)

The purpose of helping students, institutions and the general public to understand tests and their uses is sadly confounded by the provisions of this section, rather than enhanced by them. Effective decision making requires not only accuracy in the data, but appropriate *timing* in the delivery of information. For example:

Section 3.(a)(3)

This information should be provided to students at *the time test scores are released* rather than at the time of registration for the test. Students lose or discard their registration materials so frequently that testing sponsors can serve students better by providing them the information necessary for interpreting and understanding test results with the score reports when they are released. Students will be disserved by this provision in that it would require conscientious testing sponsors to provide the identical information on two different occasions, once needlessly and at added expense to students. It would be useful if proponents of the Bill were to study carefully the thoughtful means by which almost all of this section's provisions are already achieved by testing sponsors. The considerable history of the matter should not be ignored in drafting legislation.

The requirement that the testing sponsor provide institution-specific "correlation" data is inappropriate. Such data are generated only at the instruction of the institution using the tests. As such the data are provided under implied contract with the institution. There is reason to question the intent of the Congress to direct institutions to provide data of primary interest to them.

Section 2.(b)(4) urges the use of multiple criteria in those predictive activities involved in granting or denying educational benefits, while section 3.(a)(3) requires test sponsors to report correlations between test scores alone and future success in schools. If it is undesirable to use a test score as a sole criterion for admission to a postsecondary institution as suggested by the sponsors of the Bill, by what rationale does the Bill require that test agencies provide correlations between test scores *alone* and "success in college or in the career for which admission is sought"? Furthermore, the drafters of this Bill have clearly not thought through the implications of this requirement as regards post-baccalaureate admissions. Indeed, the provisions indicate a definite lack of awareness of the nature of the problems involved in both defining and measuring career success.

Congress should be reminded that a major purpose of the tests in question is typically the prediction of first-year academic performance. The developers of these tests may not wish to carry their purposes far beyond that considerably more specific scope. If a Bill is enacted that requires the information suggested by this provision, Congress will have ignored the proper uses of the examinations and, in the process, will have appeared to encourage an invalid use of some tests.

Section 3.(a)(4)(A)

The effects of compliance with this section will disappoint students and the public. We emphasize the importance of understanding that a principal purpose of the examinations in question is to predict grades for the further purpose of facilitating informed decision making of many different types (e.g., admission to postsecondary education, placement, sectioning, counseling, remediation and other educational functions). The purpose is to provide a climate in educational institutions that will enhance development of a curriculum suitable to student needs. This may require considerable intervention for some students, and more intervention at one institution than another, to enhance learning.

The Bill would require that the test sponsor provide an understandable indicator, in terms of a percentage, of the extent to which test scores add to the prediction of future grade point average over other criteria. The Bill appears to suggest that the derivation and reporting of such a statistic is a routine task for psychometricians. In fact, it is not. The sponsors of the Bill should be required to specify how such a statistic is to be derived and also to provide an acceptable rationale for that statistic.

Even if such a statistic could be generated, abstract application of this provision's requirements would add little to the understanding and intelligent use of standardized tests. It may be simple to determine and state that, on the average, the high school academic record contributes a certain amount in predicting students' academic performance in college and that test scores contribute in some other amount to the same prediction. What should be of interest to Congress and the drafters of this Bill, however, is not this abstract statistic, but that, in actual practice, the contributions of both test scores and grades *vary* across high schools. Moreover, since the standards by which high schools award grades also vary, standardized test scores become an *essential* external equalizer of the varied grading practices and curricular offerings of these thousands of secondary schools. This value cannot be expressed by a simple percentage. Indeed, to focus on this isolated mathematical relationship is to ignore one of the principal contributions of standardized testing to our educational system.

The prediction of grade point averages is only one of several useful applications of the results of these tests. Abstractions concerning the meaning and uses of such predictions will not add measurably to achieving the objectives of the proponents of this legislation.

Section 3.(a)(4)(B)

This provision appears to assume that the tests are negatively biased against students from low-income families. This conclusion is based on the well-known fact that test scores are related to level of income. It is also a well-known fact, seemingly ignored by the Bill, that high school grades also correlate highly with family income. Please refer to the table attached as Exhibit A for further information in this regard. Many other variables correlate in the same fashion with family income. Such findings are by no means evidence of bias. We do not wish to be glib, but it appears that proponents of this legislation believe they have discovered a new social phenomenon: taken in isolation, standardized test scores are directly related in some way to family income. If we are to conclude that we should discontinue using test scores, based on this knowledge, then we might also consider discontinuing all forms of college education, since it is an established fact that peoples' success in life -- at least as measured in economic terms -- is directly correlated with their educational level. Under that definition of bias, it is unfair that some persons obtain a college education and others do not.

Not only is it shortsighted to attribute bias to tests because they reveal differences in the developed abilities of students from different level income groups, it is ill-informed. Research data show that tests are effective predictors of academic performance in college for persons of all levels of socio-economic status. That is, students who score low on the tests, be they black, white, male, female, rich or poor, also tend to perform less well in college than their higher scoring counterparts.

It would be far better for Congress to direct its full attention to *why* standardized test scores (and academic performance in general) correlate with socioeconomic status, given that we know *that they do*. The latent implication that there is something wrong with the tests, that the correlation has something to do with race, is superficial. Data clearly indicate that there is as much variability on test performance within racial classification as between family income and test score performance for the population in general. That is to say, high income minorities score well on standardized tests because, like their white counterparts, they have been exposed to strong educational systems. They hold values similar to those held by high-achieving people as judged by academic criteria (which are by no means the only important criteria by which to assess our people).

Concern is sometimes expressed over the psychological effects of low test scores on the students receiving them. While sensitivity must be demonstrated in communicating such scores, surely there is no merit in altering the tests so that they minimize important differences in the developed abilities of students. To do so would be to deny students and educators factual information which they can use to plan future educational interventions aimed at facilitating access to desired educational programs and vocations.

If the Bill's sponsors are requiring that test score information be reported by level of family income to provide students from low socioeconomic status with an acceptable rationale for their lower test scores when they are attained, then they are doing a disservice to students. To suggest that the test results are irrelevant or less valid for the purposes to which they are put is not accurate. Even worse, it does nothing to address the basic challenge confronting the students -- that they are less well equipped and ready for subsequent educational experience than their higher scoring counterparts.

We acknowledge that the requirement of this section can be met. Indeed, we currently publish much of this information in some of our reports. It is critical for Congress to contemplate what extending the requirement means and what it may imply for our people. We draw no conclusions about the worth of an individual as the result of his or her performance on our tests.

Nor do we draw generalizations about his or her geneology as the result of testing. We have a *responsibility* to assist in pointing out when he or she appears not to know mathematics; when he or she does not seem to understand the English language (and, to the extent permitted by the design of our examinations, the nature of that misunderstanding); when she or he does not understand something that is held of value by our educational system.

Likewise, it is not we who determine *what* should be held of value. We leave that to educators, who are in turn accountable to legislators and ultimately to the public. When a curriculum is determined, we are capable of developing instruments that, when properly interpreted, add measurably and provably to our understanding of what steps an individual should take in continuing his or her education. *We cannot say why there may be deficiencies in the preparation of an individual for onward educational experiences. To focus on our instruments with this expectation, to blame them for these voids in preparation of some of our people, is to profoundly ignore our most pressing social problem: we are educating our young people unevenly. We have yet to address this problem cogently. The problem was left to the states by the U.S. Supreme Court in* _*Serrano vs. Priest*_. To the extent that this problem is properly one for the federal government, it would seem better for Congress to turn its attention to the *reasons* for the disparities in preparing our youth than to assail tests because they call attention to the disparities.

Section 3.(a)(4)(C)

Compliance with this section would mandate research of a particular character by testing sponsors. We question the wisdom of any legislative mandate for research by Congress, least of all research on the coaching topic. Logic dictates that "test preparation courses" that are of sufficient length and content to be considered *instruction* need little investigation, since their worth is obvious. They will help some students and they won't help others, as is the case with any instructional program. To the extent that these "courses" don't qualify as instructional in nature, study of them by testing sponsors requires the sponsors to validate or invalidate the efforts of organizations whose social contribution they sometimes believe to be questionable and whose procedures frequently conflict with the sponsors' test security measures. Given the patent conflict of interest involved, it appears that Congress would be well advised to secure more suitable avenues for both developing and disseminating the information required by this section.

We point out that the recent FTC study of test preparation courses lead the Commissioners to refrain from providing specific guidance to students and testing agencies regarding preparatory courses. H.R. 4949 goes beyond the FTC's actions and findings under the assumption that accurate, helpful information is known. Telling students that "some preparatory courses may be helpful to some students for some tests" is hardly helpful, but it is the only reasonable statement that can be made given present knowledge of the topic. To state otherwise would be misleading.

Section 3.(a)(7)

The definition of "property interests" is a matter for the courts. Regardless of the nature of any statements made by testing sponsors on "property rights," some will disagree and seek judicial judgment.

The rights of students with respect to the disposition of the information they provide ACT are described in considerable detail in Exhibit B* at the end of this document (see "Statement of Policies -- Use of ACT Data, Privacy, Responsibilities to Program Participants," attached). These policies are rigidly adhered to. They predate any legislative initiative or expression of concern by student advocates. As can be seen from our comprehensive policy statements on the topics covered by this section (twenty-three pages), the task of communicating and fully interpreting these policies in any meaningful detail to over a million students (or within the contents of the 2,700,000 registration packets we distribute annually) would be incredible but could be done. Because the benefits of such an ill-conceived action would be almost nil, we advise against doing so. Much of the requirement of this section is already attended to by our materials and those of other testing sponsors.

*Retained in Subcommittee files

We believe that Congress should be mindful of what is already being done and with what conviction by testing sponsors. The legislative requirement in this section adds little to an already thoughtful and conscientiously-administered program of policy development, procedural implementation and communication.

Other testing sponsors have and adhere to such policies, concomitant with their substantial responsibility to the American public, and within the purview of extensive governance by bodies accountable to the public.

Section 3.(a)(8)

Much is said in Section 3. about describing to students any promises or covenants made to them by the testing sponsors. Whether the purpose of this requirement is to suggest that there is a "contract" between the testing sponsors and the test subjects, or to suggest that testing sponsors are casual in matters that could result in delayed score reporting, the legislative "solution" posed in this and related sections (e.g., Section 3.(c)) is totally without merit. The reporting and system requirements involved subsequent to release of test scores are extensive and hinge entirely on timely and effective conduct of the processes of test administration, scoring and reporting. It is totally unwarranted to assume that there is *anything* to be gained by testing sponsors in reporting test scores late, incorrectly, or in a haphazard manner. Practically speaking, the Bill will invite testing sponsors to inflate the reporting period stipulated in this provision to avoid incurring the useless reporting requirements.

Experience with testing systems of the size and complexity involved with undergraduate admissions indicates that late test score results (which are extremely rare) would reach students and institutions about as quickly as the notice this and related sections would require. We assure Congress that operational and business considerations, not to mention the totally wasteful additional costs to students that would be incurred in implementing this provision, are sufficient to motivate testing sponsors to perform their services to students and institutions efficiently and effectively. With the added consideration that the testing sponsors *believe* that their primary functions are services to the public, with particular emphasis on the needs of students and educational institutions, and in cognizance of those needs, this requirement is without any redeeming value.

Section 4.(a)(1)(A)

This section implies that testing sponsors conduct many research studies that they do not make public out of self-interest. Such an implication is totally without substance. It is true that certain studies compiled annually for thousands of educational institutions, for example, studies that provide a profile of the students at individual institutions, are maintained confidentially -- for good reason. There is considerable risk in requiring testing sponsors to deposit such studies with the Commissioner of Education or any other office or agency. *Incorrect and often pejorative* conclusions about the quality of these institutions will be drawn by many based on the data. Inappropriate comparisons across institutions are inevitable. It is equally inevitable that many institutions will discontinue such studies rather than face the gauntlet of ill-informed public conjecture. The result will likely be *less intensive* public scrutiny of the uses and validities of those uses for the examinations. The irony is anguishing.

Pertinent studies of the nature of the tests covered by this legislation and their statistical characteristics are now, and have been for many years, freely available to the public. There is no need for a legislative dictate with respect to these data. On a minor note, Congress should be mindful of the sheer magnitude of the reports that will potentially flood the U.S. Office of Education if this requirement is not changed. The reports could easily number in the thousands each year for all testing sponsors combined. Unless a plan existed for the constructive analysis and use of the data in the reports, it is difficult to see on what basis Congress could justify requiring their submission.

Section 4.(a)(1)(B)

This section conflicts with existing federal law on freedom of information (protection of trade secrets) and restraint of trade (stifling competition through sharing of cost/pricing data). Many of the contracts covered by this section result from a *competitive* bidding process. Congress must give careful thought to the subverting impact the release of these contracts may have on that competitive activity.

Section 4.(b)

The purpose of this section is unclear. The focus on comparisons between the variables identified and test scores alone touches again on symptoms rather than the causes of the educational inequities in this country. The mere relationship between test scores and the variables cited is of considerably less significance than the cause of this relationship, something which the tests do not address. *Whatever* the relationships between test scores and the variables cited, similar relationships will be found between these variables and other academic assessments, thus confounding any causal interpretation.

With respect to the report concerning the relationship between test scores and the completion of test preparation courses, it should be made clear that this requirement is defective. The extant data on this topic do not constitute an adequate inquiry. If the information is to be provided, Congress would be well advised to provide resources and a charge to one of the many federal agencies (e.g., NIE or NSF) qualified to conduct or sponsor such research to proceed with the design of studies aimed at shedding light on the issue in all of its practical complexities.

Section 5.(a)

This Subcommittee, and state legislatures considering similar legislation, have heard much about the public release of "secure" standardized tests. We cannot add greatly to what has already been said. We summarize here several issues, relating the apparent objectives of test release proponents to the counterview.

"Students can monitor test scoring accuracy" by viewing questions and their corresponding correct responses. This is uncontestable. It is a simple task because the precision with which the tests are constructed requires avoiding uncertainties and ambiguities wherever humanly possible. The result is that automated scoring from a key is not a complex task. The "manual" verification of this process by millions of human beings is the only advantage we can see stemming from this requirement. Its necessity, however, is gravely questionable. Other reasonable alternatives exist and are, in fact, in place.

"The public can scrutinize the contents of examinations" to ascertain qualitative information, and to examine potential for bias and discrimination (in the social as opposed to the psychometric sense). The tests are developed with the assistance of experts in curricula from all over the United States. Development involves minority review. Moreover, tests are already made available to qualified researchers who have the tools and the professional capacity to make such evaluations. Errors in test items of the nature implied by this objective are rare. The testing agencies are the first to react when they are noted, as they must. It should be understood that absolutely no item can be constructed that won't find at least one disenchanted reviewer. The tongue-in-cheek mathematical contest of 2 + 2 = 4 presents suitable fodder for challenging the inclusion of that item in any kind of assessment, if one wants to indulge in suitably-prolonged philosophical inquiry. Carried to an illogical extreme, the testing sponsors could not successfully construct any kind of examination if it had to be designed by a committee of the whole, all of whom would concur without reservation on the nature and content of the items involved.

While there is truly nothing to hide from members of the public in the content of the examinations (they see them at testing time, and faculties can and frequently do review them before deciding they are appropriate for their uses), there is much to lose by saddling testing sponsors with responding to myriad inquiries regarding an item or items about which there is absolutely nothing wrong. Those that might be wrong, however rare, are righted by test makers and test subjects in the normal process of test design, construction, administration and use. The costs to students for the test sponsors to respond to potentially thousands of letters and questions about test items must also be considered in light of the ultimate benefit of that process.

"Students can learn" from seeing what they missed on an examination and knowing the correct answer(s). This is undoubtedly the case. The important question is: how much can they learn and to what end? Very little, and to no good purpose. Most tests covered by this legislation sample academic disciplines and domains only to the extent appropriate to the uses of the test(s) in question. These tests are not sufficiently comprehensive to allow for in-depth diagnosis of educational shortcomings because they are not designed for that purpose. For a student or teacher to conclude that because he or she missed two or three algebra items, she or he doesn't "know algebra" would be distressing. One cannot conclude from any rational exploration of this objective that disclosure of these tests, tests that are not designed for and *should not be used in a direct pedagogic manner*, will in any truly meaningful way provide a learning experience for test subjects. To conclude that these particular tests would provide a learning experience is to engender test abuse beyond that alleged by some proponents of this legislation.

We can state a number of potentially enormous disadvantages that will result from the release of these tests to the public:

Teachers will be tempted to teach for the tests and many will succumb to the temptation because they will feel duty-bound to provide their students with the best opportunity of performing well. They will also be tempted to teach for the tests to demonstrate the success of their own efforts in behalf of the students.

Teaching for the specific items included in a test perverts the purposes and the effective uses of the test. We will not further elaborate those purposes and uses at this point because of their complexity. They should, however, be of interest to Congress and to everyone concerned with education. *The issue involved is crucial.*

The validity(ies) of the tests released will likely be lowered over time. This expectation is partly attributable to the tendency of teachers to teach for the tests, and for the tests to control the curriculum as a result of that activity. The curriculum should dictate the content of the tests rather than the reverse. The problem is that while increased test performance might result from teaching for the tests, the postsecondary curriculum will remain defined apart from the secondary school teaching activities. If inappropriately "inflated" test scores are used in predicting first-year college performance, then the validity of the decisions made using those predictions will likely not be of the quality presently enjoyed. If the objective of releasing secure tests and test items is to destroy the contributions of the tests, then this requirement has merit. If that is not the objective, if instead the objective is to improve the usefulness of the tests, this requirement will be counterproductive.

Costs will increase. This requires little elaboration. It should not be the sole basis for determining the benefits to be derived from the mandated release of secure test materials.

Placement of college students in courses will be less effective than at present. If disclosure of the test materials results in less valid instruments, the test data will have less value for course placement. This could cause a disruption -- to some extent unascertainable at this time -- of the placement procedures and instructional programs at some colleges and universities.

The educationally-advantaged will enjoy even greater advantage over their less-advantaged peers. The better, economically strong, school districts will have, as they always have in the past, greater success in teaching students the knowledge and skills that effect test score performance than those schools with fewer resources and other non-educational distractions. Some argue, to justify this legislation, that schools with strong economic bases presently teach for the tests, while those with fewer resources do not. If this is indeed true, it does not mean that releasing the tests will result in eventual equity in balancing this teaching effort. Rather, it is more likely that the schools with more resources will do an even better job of teaching skills pertinent to the tests, while the less-advantaged schools will continue to struggle with the more fundamental problems they face. The test performance engendered by the schools has nothing to do with the availability of the tests. Rather, it has to do with scarce resources, parental and peer pressure, students' aspiration levels, and related factors, as mentioned in our comments above concerning the implications of Serrano vs. Priest. The effect may be to exaggerate differences between the rich and the poor, contrary to the supposed intent of this legislation.

Section 6.

The term "specifically authorized" as used in this section will deny students and institutions valuable services by some testing sponsors. Test subjects are currently offered an opportunity to authorize testing sponsors to communicate with them about educational experiences that may be of interest to them, or to authorize the release of their test records to institutions and scholarship programs on a non-specific basis. Most students elect this option, and must do so before their records may be so released.

Students receiving information in this manner have been pleased with the results. Institutions and scholarship agencies are able to inform minorities, handicapped students, and students with special learning disadvantages of their educational offerings. Requirement for "specific" authorization obviates the provision of such services by the testing sponsors.

The final provision of Section 6. restricts release of test records (unidentified) for research purposes primarily concerning the tests themselves. This restriction will unnecessarily deprive researchers of a valuable source of data related to a variety of important topics. This section should be modified to allow research with test results so long as identifying information is expunged prior to release of the data.

It should be noted that this section conflicts with both the spirit and the letter of recent recommendations of the Privacy Protection Study Commission. We suggest a review of the Commission's recommendations and the underlying rationale for release of research data before Congress acts further on this provision.

Section 7.

We do not object to fiscal disclosure. Our fiscal information is already a matter of public record. It should be understood that the form in which it is disclosed has implications for competition with other agencies. Further, we believe no conclusion should be drawn about the operation of a testing program in isolation from the overall plan and developmental activities of the testing sponsor.

Sections 8. and 10.

It is likely that few, if any, testing sponsors can meet the Bill's requirements within the contemplated effective date without curtailing services. This is a simple fact. It will take up to two years before compliance with the requirement to publish test contents can be met, and for some sponsors compliance may be impossible regardless of the time frame for implementation.

Summary

We have provided the foregoing comments on H.R. 3564 and H.R. 4949 as part of what we believe to be our responsibility to American education. It is our conviction that the proponents of this legislation, while well-intentioned, have concocted a remedy whose ultimate consequences will likely overshadow the illness supposedly being addressed. We are concerned that the *very* important reach of these bills within our educational system -- both in terms of pedagogy and in terms of the proper role of government -- has been given short shrift.

It is crucial that subsequent study of the bills by the Congress emphasize a thoughtful assessment of the *need* for this legislation balanced against the impact it could have. The potential negative consequences, to some extent motivated by lack of understanding of the nature and purpose of the tests themselves, surely deserves more serious attention than these hearings will provide. To that end, we strongly urge thorough subsequent study of the testimonies provided.

We are dealing with the potential of literally millions of persons. It is undeniable that tests play an important role in this context. That role is appropriate in the educational setting and has been found sound in the occupational setting by the U.S. Supreme Court. This legislation jeopardizes the contribution of tests to our educational system and ultimately poses grave problems for the autonomy of our institutions of higher learning. The burden of proof regarding the need for legislation on standardized testing cannot be met by political sloganism or populist concepts put forth without benefit of data that verify the nature and extent of the harm proponents of these bills assert applies. Until such data are provided, the proposed legislation should not be enacted.

Family Income, High School Grades and ACT Test Scores*

	Under $3,000	$3,000-5,999	$6,000-7,499	$7,500-8,999	$9,000-11,999	$12,000-14,999	$15,000-19,999	$20,000 and over
Average ACT Composite Score	13.5	15.2	16.9	17.6	18.6	19.4	19.9	20.6
Average High School Grade	2.67	2.82	2.95	2.97	3.01	3.05	3.08	3.07

---

*College Student Profiles: Norms for the ACT Assessment, 1978-79 Edition. Based on a random, selected 10% sample of 476,822 students who took the ACT tests in 1976-77, and who subsequently enrolled as freshmen in one of the 1,138 colleges who participated in ACT's Class Profile Service.

EXHIBIT A

Chairman **Perkins**. Thank you very much.

We will go along to Dr. Perrone.

Without objection, all the prepared statements will be inserted into the record.

STATEMENT OF VITO PERRONE, DEAN, SCHOOL OF EDUCATION, UNIVERSITY OF NORTH DAKOTA

Dr. **Perrone**. Mr. Chairman, members of the subcommittee, I am Vito Perrone, dean, Center for Teaching and Learning, University of North Dakota.

I am pleased to have this opportunity to address the House Subcommittee on Elementary, Secondary, and Vocational Education in regard to the proposed Truth in Testing Act of 1979, H.R. 3564, and the Educational Testing Act of 1979, H.R. 4949.

My testimony grows out of my experience over the past two decades as a student of testing and an active participant in the national dialog over tests and test use.

In my roles over the years as an elementary and secondary school teacher, university professor, graduate dean, dean of a professional school, coordinator of a major national study group on evaluation and President of the National Consortium on Testing, I have had numerous opportunities to observe the influence of externally organized tests on the lives of large numbers of individuals.

The Study Group and the National Consortium on Testing are organizations committed to such goals imbedded within the proposed legislation as increasing public understanding of the strengths and weaknesses of standardized tests and improving the quality of testing and assessment practices.

In my work, I have been privileged with opportunities to interact regularly with creative scholars who are concerned about assessment practices. I have also had numerous occasions to meet with officials of our major testing companies. Having said who I am and something about my experience as well as having identified several affiliations, I need to say that my perspective is not to be viewed as an official position of the University of North Dakota, the North Dakota Study Group on Evaluation, or the National Consortium on Testing.

Standardized testing, essentially a post-World War I phenomenon, has become a way of life in America. Tests exist for almost every human social trait imaginable from intelligence to alienation, self concept to maturity, moral development to creativity. They are being used to select people into and out of a wide range of educational programs, private and public projects, and a variety of jobs.

We have become so accustomed to their use and their seeming scientific authority that we have often failed to ask whether the tests do in fact assess what they purport to assess or whether the assumptions that undergird the statistical and psychometric constructs used are acceptable.

Increasingly, however, unexamined acceptance of testing practices and use is being replaced by critical scrutiny. To some degree this has occurred in the wake of an explosion over the past two decades in test use and in the growth of the testing industry.

Individuals as well as organizations are becoming increasingly concerned about reports of misuse of test results.

They are growing in their concerns also about the limited public information available about tests and test use. The two legislative proposals that are before this committee, as well as similar legislation passed into law in New York and being prepared for legislative action in a number of other States, are reflective of these growing concerns.

I believe that opponents as well as proponents of the proposed legislation believe philosophically that those who take tests and use the results of tests need access to a wide range of information such as what the tests are about, what they purport to measure, scoring mechanisms, timeliness of score reporting, data concerning reliability and validity, et cetera. Where they differ, I suspect, relates to how much information needs to be made available, the form of the information and the accessibility of the information.

The legislation before this committee, principally sections 3 and 5 of H.R. 4949 and to a certain degree section 6 of H.R. 3564, calls for all information to be a matter of public record rather than that which a test producer chooses to make available. Public accountability in the largest sense is what the legislation demands. Given the role that tests play in American life, such a demand is not unreasonable.

I should note early in this testimony that my remarks will focus, for the most part, upon H.R. 4949, inasmuch as it addresses exclusively an area in which I have the greatest amount of experience. H.R. 3564 is less comprehensive in regard to reporting requirements but adds, in contrast to H.R. 4949, concerns about occupational testing.

This is an important issue that the courts have been involved with for some time. That greater public accountability for occupational tests is needed is unquestioned. While focusing on H.R. 4949, I wish to acknowledge that the bulk of my arguments about openness relate also to the major provisions of H.R. 3564.

Sections 3, 4, and 5 of H.R. 4949 which would require relatively complete disclosure of all information relating to an admissions test and its content and upon request return of corrected test questions to individual test takers go to the heart of my concerns. I believe they are the critical sections for opponents as well.

Section 5, in particular, comes out unequivocally for the principle of fairness; that those who are judged by test results are entitled to know the basis for whatever judgments are made. To be able to examine their actual responses in relation to the items that made up the original tests is hardly unreasonable.

In its editorial of July 24, 1979, in response to the passage of the New York truth in testing bill, the New York Times wrote in a straightforward manner:

> In spite of some concerns there is the matter of elementary fairness. These academic tests help to shape the course of people's lives, their schooling, their careers, the very sense of their own abilities. Students deserve to know how they are being rated and judged. There are freedom of information laws, truth-in-lending laws, truth-in-packaging laws. Why not truth in testing as well?

One of the arguments originally advanced by opponents of more openness in testing was that release of items used for equating tests across forms and years, for purposes of determining comparability, as well as for pretesting process, for purposes of developing new test items. H.R. 4949 recognizes this concern and allows for maintaining the security of items which do not contribute directly to an individual test taker's raw scores.

This acknowledgement in H.R. 4949 ought to assure opponents of the legislation that the intent is not to bring an end to standardized tests. The legislation acknowledges the important public policy role that standardized tests play and seeks improvement and increased fairness, not demise.

Test publishers have claimed, in response to calls for openness, that the technical qualities of their tests, including validity, reliability and item quality, will suffer irreparably if the actual tests are released.

Such critical qualities as validity and reliability should not be affected negatively if test publishers bring to the task of openness the same degree of knowledge and commitment they bring currently to the promotion of secure tests.

By making the tests and background about the tests publicly available, the quality of the public debate should certainly be enhanced and the quality of the tests improved. At the moment, test items actually appearing on the tests are not subjected to the scrutiny of a broad range of subject matter experts, educators and citizens with a view toward improving them.

Reviews of items made available on occasion have not been particularly positive, and in my presentation to this committee, I have included some particular examples of the impreciseness, the ambiguity, and the single logic that seems to be accepted.

[EDITOR'S NOTE.—The paragraphs in the written statement which were omitted in oral testimony follow:]

Banesh Hoffman's analysis of SAT items in "Tyranny of Testing" (1962) is only one classic example. And reviews of other standardized tests which use the same technology incorporated by the principal developers of admission tests (ETS and ACT) do not encourage high levels of confidence about existing item quality or validity. For example, critics note that the multiple choice formats necessarily provide ambiguities, the possibility, even if well written, for particular individuals to argue for more than one of the answers. The problem is, however, that only one logic is permitted. Let me provide an example cited by Banesh Hoffman (New York Times Magazine, 24 October 1965) and elaborated upon by Judah Schwartz (National Elementary Principle, March/April 1975). How should one respond to the following verbal analogy: Island: Water. (a) Sand: Desert. (b) Mountain: Ocean. (c) City: Meadows. (d) Lake: Land. (e) River: Banks.

It is possible to argue a logic for each choice and I will do this if the Committee members are interested. What is clear, however, is that "the quality of any answer offered to a verbal analogy can only be judged in the light of the respondent's stated rationale for his or her answer." (Schwartz). Even information items, depending on how deeply an individual thinks about the potential responses, may be less straightforward than a surface examination would suggest.

Dr. PERRONE. The test publishers and their client groups argue that item quality and validity will necessarily decline if the legislation is enacted because the potential pool of items for use in the tests is limited. From what it is possible to learn, and it is not easy now to get much information, very few items are reused; hence, every form of most admissions tests makes substantial use of new items.

The potential size of the item pool has not yet been a significant problem and is not likely to be an insurmountable obstacle if the proposed legislation goes into effect. It should be clear that in most areas covered by admissions tests, the potential item pool is almost infinite if item-generating algorithms are developed, especially if test publishers examine seriously the potential of test constructs other than the current multiple choice, single response format. If the public disclosure requirements cause test makers to consider alternative constructs, another useful end will have been served.

In relation to the foregoing issues, it should be noted that we have a history of public disclosure of tests. External certifying tests used in the English schools were made available to the public soon after administration. The Regents examinations in New York have had a similar history. And publicly released items from the national assessment program are currently used in a variety of State testing programs.

Few have argued, to the best of my knowledge, that the quality of these tests has suffered from the public disclosure. Rather, the concern has been that public disclosure has encouraged teaching to the tests. Has teaching to the tests been absent in a milieu of secure tests? Obviously not.

Test coaching has become a major new industry in the United States. While controversy exists about whether test coaching schools are effective, the consumers clearly believe they are. Those who carried out the FTC study in Boston are also convinced that the coaching is effective for those whose economic circumstances permit an expenditure of $300 or more.

The openness called for in the legislation will enlarge opportunities for all students to have equal access to testing information and, if they have taken a particular test, to learn from their mistakes.

If students learning nothing from a testing process which takes their time and money, a problem surely exists. Teachers within schools and within cultural and community settings might also learn enough through the review of accessible public information to assist all students in their preparation. Openness, then, has the potential for standardizing to a larger degree than now exists the conditions under which individuals take tests.

I realize that the foregoing is one of the arguments now used by opponents of the legislation, that teaching to the tests will be encouraged and test scores will as a result have less credibility. The fact is that teaching to the tests already exists but more differentially than concerns about equity should support.

It could, of course, be further argued that public discussion would bring about greater attention to the importance of particular skills. If, as a result, "Teachers teach their students how to carry out this particular task, so much the better; education will have been improved in the process" (Edwin Taylor, EDC News, Winter 1978).

The test scores in the new circumstances might, as opponents of the proposed legislation suggest, result in lower levels of credibility. This, however, does not suggest uselessness. A lower level of credibility might, in fact, encourage those who make admissions decisions to use a broader range of criteria, a purpose outlined positively in H.R. 4949, section 2, paragraph 4. Even the testmakers stress increasingly that a student's performance on a test is only one and possibly not the most critical indicator of their ability to succeed in a particular college or professional program.

Attached to my formal testimony is a paper which I prepared a year ago on college admission tests and provides, I believe, an important perspective on the foregoing issues.

Opponents of the proposed legislation argue that the costs for such tests as the SAT, MCAT, LSAT are likely to treble if the legislation is passed, thus working a significant hardship on many students. Information presented by ETS at the May 9 hearing on the New York truth-in-testing bill hardly supports such a claim.

Andrew Strenio, in a paper prepared for the National Consortium on Testing entitled "Open vs. Secure Testing," which will be entered into the subcommittee record prior to October 8, quotes from several ETS documents which show that actual test development costs range from 4 to 8 percent only of the total cost of a major test.

> In 1971-72, this amounted to 32 cents of the $5.75 paid by an individual taking the SAT and 62 cents of the $8 paid by an individual taking the GRE.

This cost is small. Admittedly, returning all test materials to a test taker would create some additional costs beyond those related to test development. These increases, however, should not be excessive in relation to the potential gains and should not be difficult to adminster. Acknowledging this possibility, H.R. 4949 in section 7 offers support for testmakers to add the actual costs of providing such services to students.

Many of the benefits of openness have been discussed. I will add to the list in order to give greater credence to this position. At the level of the consumer—the individual taking a test—enormous anxieties tend to exist. They understand that their lives will be influenced significantly without knowing very much about the sources of the influence, feeling inadequate to control this segment of their lives.

They receive scores in the mail but are left to wonder what particular questions or types of questions they missed or got correct; whether the scoring was accurate or inaccurate; whether the scoring procedures were consistent with other test administrations, and the like.

This, it seems to me, is an intolerable condition, not likely to be tenable in most other educational or employment settings. Are mistakes ever made? The major testing companies have, without question, established careful procedures for assuring the accuracy of their scoring and reporting. Nonetheless, there have been reports of serious errors or misinformation during the past 3 years.

For example, ETS changed the content of the LSAT between the July and October 1977 administrations without reporting the difference to law schools. The result was that students who took the test in October had, on the whole, higher scores than students who took the test in July (New York Times, April 28, 1978).

In May 1978, ACT, which administers the MCAT, wrote to medical schools to inform them that medical school scores they had sent were incorrect (Boston Globe, May 23, 1978). During the 1975-76 year, ETS acknowledged that it had erroneously designated some law school students as "unacknowledged repeaters" (New Republic, March 12, 1977). How many other errors have there been? How have they affected individuals?

How much information about tests and testing is available for public consumption now? Are the tests fair in a substantive sense? Do they ask important questions? Do they favor particular modes of thought over other modes? Are they culturally or linguistically biased? What underlying data exist to support existing validity claims or fairness claims?

The difficulty in responding authoritatively to these kinds of questions is that information is not easily attainable.

[EDITOR'S NOTE.—The paragraphs in the written statement which were omitted in oral testimony follow:]

> While testing enterprises argue that they make information available to "qualified scholars," there is ample evidence that an information gap exists. Oscar Buros, who has been the principal chronicle evaluator of tests over the past forty years, declared in his valedictory address (Educational Researcher, July/August 1977) that "the information available to permit an adequate assessment to be made of (test makers) secure tests is quite unsatisfactory. Although our reviewers generally receive some in-house material, which is not available to other educators and psychologists, even this material is inadequate."

While there is a belief, from my perspective, substantiated to a large degree among many groups such as the Association of Black Psychologists, the NAACP, Mexican-American Defense League, the Southern Regional Council, and the National Council of Teachers of English, among others, that current tests are subject to considerable cultural and linguistic bias that compromises the country's commitments to educational equity, public accessibility to tests, test scores, validation, demographic and ethnic data, et cetera, may provide the substantive information that is so critical to addressing these concerns thoughtfully and authoritatively.

Section 6 of H.R. 4949 addresses the concern for privacy of test scores. This is a sensitive statement that grows out of considerable confusion about whether privacy is currently honored. Many student groups are currently challenging ETS on this issue. My sense is that no one opposes seriously this basic provision of the proposed legislation.

In closing this statement, I wish to summarize the principal benefits which can accrue from the proposed legislation: Increased levels of public information about standardized tests; public accountability of those who produce tests; fairness to those who take tests; improved quality of test items and validity; encouragement of alternative test constructs; greater equity in relation to accessibility to test taking information; more authoritative responses to critical questions about testing and test use that have plagued us for a half a century, essentially questions such as cultural and linguistic bias; improved admission procedures that concern themselves with a broader range of criteria; greater assurance of privacy.

I am aware that requiring public disclosure will not resolve all of the problems of equity or assure in any absolute sense that the principle of fairness is guaranteed. But the requirement is clearly a step in those directions.

I encourage this committee to support the proposed legislation.

[The appendix attached to the prepared statement of Vito Perrone follows:]

Appendix to the Prepared Statement of Vito Perrone

College Admissions: A Perspective Presented to the 1978 National Forum of the College Board, Vito Perrone Center for Teaching and Learning, University of North Dakota, October 30, 1978

I am pleased to be here for this particular discussion on the "Role of Standardized Tests in College Admissions, Credit and Placement" programs. While I believe that standardized testing efforts at the elementary school level—especially in the very early years—have a more pernicious quality, college level testing remains of great concern to me.

You should know at the outset, if you don't know already, that I am not an admirer of the standardized testing practices that have become so commonplace in our country. I remain unconvinced that the gains which have accrued—however one wishes to define them—outweigh in any significant manner the losses. I believe we have concentrated considerable energy and fiscal resources on a process having enormous limitations without making sufficiently manifest those limitations. We have constructed a mechanism and related support systems clothed in the language and mystique of science with restricted accessibility being a major by-product. Testing has become so much a part of our educational structures—the conventional wisdom—that its purposes do not receive sufficient challenge and any significant examination of how well those purposes match use is left wanting.[1]

I am aware that we are in a major "standards debate" today in which test scores are playing the role of Hermes. This 70's drama, unfortunately, tends to restrict our ability to address critical issues in testing at a time when we can countenance it least. My hope is that we can, at least for the sake of this particular discussion, set Hermes' message aside.[2]

At the heart of much of our post World War II debate about testing in relation to college admissions is the classic tension in American society between egalitarianism and meritocracy, alternative definitions of democracy in practice. Whether that tension surfaces forceably or remains obscured is determined largely by a wide range of social, economic and political circumstances. In periods of stagnation and uncertainty, times of crisis, which may well describe the present, arguments which tend to support meritocracy become particularly visible.[3]

It might be useful to place before you an historical perspective that may assist in understanding some of the present concerns about testing and college admissions. (This perspective will suffer somewhat from its summary nature but that is the dilemma one faces in this kind of presentation.) Prior to World War II, testing had a relatively benign character in relation to college admissions. Very few young people were completing high school and of that population only a small percentage considered entry into a college or university. At the turn of the century, for example, fewer than 2 percent of any age cohort completed high school. This number rose to approximately 20 percent in the early 1920's, reached almost 40 percent by the beginning of World War II and in my high school graduation year—1950—hit 52 percent. For much of that half century period individuals were not typically in a disastrous economic or social position by virtue of a schooling deficit. (This is not to suggest, of course, that maximum opportunities existed for those who had not completed secondary school.) As high school attainment increased by 1965 to its present level of 76 percent, paralleled by an unprecedented growth in the population and in the technological nature of our economic system, not having a high school diploma became a disaster of immense proportions. That the disaster has befallen the poor and the minorities almost exclusively exacerbates the consequences.

---

[1] I have not attempted here to outline the technical problems which exist or the assumptions which tend to undergird standardized testing efforts. The following publications might be useful for those who are not yet involved in the literature: Vito Perrone, Abuses of Standardized Testing, Phi Delta Kappa Foundation, 1977; Testing and Evaluation: New Views, Association for Childhood Education, 1975; Henry Dyer, "Testing Little Children: Some Old Problems in New Settings, Childhood Education, /April 1973; Edmund Farrell, "The Vise/Vice of Standardized Testing: National Depreciation by Quantifiation," National Council of Teachers of English, Spring 1976; Leon Kamin, The Science of Politics of I.Q., John Wiley, 1974; Murray Levine, "The Academic Achievement Test: Its Historical Context and Social Functions, American Psychologist, March 1976; Report on Minority Testing, NAACP, May 1976; Barbara Zasloff, A Bibliography on Bias in Intelligence and Achievement Testing of Children and Youth, N.D. Study Group on Evaluation, 1975; Banesh Hoffman, The Tyranny of Testing, Collier, 1964; Deborah Meier, Reading Failure and the Tests, Workshop Center for Open Ed, 1973; National Elementary Principal, March/April 1975 and August 1975.

[2] The debate about how to interpret test score declines has generated considerable verbage but not much enlightenment. Competency testing activities which are generating considerable enthusiasm in a number of state legislatures and state departments of public instruction represent one response. I am not impressed by this "movement." It is fraught with a wide range of technical, pedogogical and social problems. College Board participants might find the following papers helpful in understanding competency testing activities: Walter Haney and George Madaus, "Making Sense of the Competency Testing Movement," Harvard Ed. Review, Winter, 1978; Gene Glass, Minimum Competence and Incompetence in Florida," Phi Delta Kappan, May 1978; Merril McClung, "Competency Testing: Potential for Discrimination," Clearing House Review, Sept. 1977; A Citizen's Introduction to Minimum Competence Programs for Students, S.E. Public Ed. Program of the American Friends, Feb. 1978.

[3] The popular press has, over the past year, carried a variety of features on the "revolt of the middle classes" which in many ways epitomizes this point.

As high school attainment has enlarged, approaching a long time egalitarian goal, the goal has escalated; post secondary education is now viewed as the minimally acceptable level of education. Given the democratic ethos that fills our language, if not always our social and public policy, could we have expected anything less? Prior to World War II, those who went on to college tended to be individuals with financial resources—essentially persons from middle and upper class backgrounds—and a desire to enter a professional field which demanded college preparation. As most of you must know, the list of fields was relatively small to at least 1940; certainly it didn't include most areas of business, the college boom field of the 1970's. Only 10–12 percent of 18–21 year olds were in college in 1940. By 1972, in contrast, the number had reached approximately 50 percent. In the wake of the burgeoning numbers seeking college preparation of one kind or another, testing ceased its benign role, becoming in the post World War II period a principal selection vehicle for our major institutions, one which fit many individuals' beliefs in "meritocracy in practice."

The theory accompanying the increased power of an aptitude or achievement test for purposes of selection was that family background and financial resources should not continue to be the principal basis upon which individuals gained access to the best of higher education's resources. Testing in this sense was viewed as a social equalizer. The theory, even as scholarship funds enlarged, never quite worked in practice. The most selective of our colleges and universities had populations after World War II that were not radically dissimilar from their pre-WWII populations. Sam Kendrick raised this issue in 1967, suggesting that too heavy a reliance on the SAT, a condition he noted, would work against diversity. He wrote, "not more than 15 percent and perhaps as few as 10 percent of Negro high school seniors would score 40 percent or more on the verbal section of the SAT. Only 1-2 percent would be likely to score 50 percent or more."[4] Now many schools by 1970 were beginning to make less use of test scores but the belief has persisted, nonetheless, that test scores are the principal determiners, especially for the poor and the minorities.[5]

In order to accommodate the growth of applicants in the postwar period, many colleges and universities expanded and new institutions were established. When I began my undergraduate study at Michigan State in 1950 there were 15,000 students. By 1960 the number was approaching 25,000 and by 1970, 40,000. And that configuration of growth was matched by colleges and universities in almost all of our States. While SAT and ACT scores during this period began to play a large role in the more selective of our colleges and determined even in some of our less selective colleges the level of enriched opportunities that would become available, they clearly were not fixed barriers to college attendance—especially if one wanted to remain close to home.

But as college attendance has become more commonplace, especially in a sputtering economy that no longer absorbs graduates in the manner that was true in the 50's, 60's and early 70's, the particular college one attends is enlarging in consequence. In the boom period we established willy-nilly a higher education tracking system.[6] Our upper track institutions tend to be better supported than lower track

---

[4] Sam Kendrick, "The Coming Segregation in Our Selective Colleges," College Board Review, Winter 1967.

[5] Harvard, recognizing the effect of the SAT on its selection process, began moderating its value after 1960, giving in the process more weight to what was called "Personal ratings." The class of 1964 was, as a result, broader in most respects from earlier classes. (See Dean Whitla, "Evaluation of Decision-Making: A Study of College Admissions," in Handbook of Measurement and Assessment in Behavioral Sciences, ed. by Dean Whitla, Addison Wesley: Reading, Massachusetts, 1968.) Harvard has, since 1964, continued to make less use of the SAT as a significant factor in admissions. While I haven't carried out any systematic review of practices everywhere, it seems clear that many institutions have moderated their use of admissions examinations. That they insist on the test score being presented, however, helps keep up the illusion, if not the fact, that it is somehow critical and that a below median score, regardless of other circumstances, might prevent entry.

[6] Jerome Karabel is beginning to address the effects of this tracking system, especially as it relates to minorities and poor. See: "Community Colleges and Social Stratification," Harvard Ed

Footnotes continued on next page

institutions—they tend also to have a much lower percentage of minorities and the poor within their student populations. They are more selective, tending to give greater weight or the illusion of greater weight to the SAT or ACT in the admissions process. Increasingly, employment opportunities for graduates of the upper track institutions have expanded in relation to graduates from lower track institutions and graduate and professional schools—which open up yet additional opportunities—are more accessible. So much for a brief historical account. It may have been a long route to the issues that I wish to raise but I believe the path was critical.

What happens when test scores are made the major factor upon which admissions decisions are made? Those selected tend to be white, middle and upper class persons from families endowed with college and post graduate degree holders. If high school grades alone are used, those selected will come from a broader socio-economic background, with more minorities represented. Now most selective admissions institutions insist that they use both test scores and high school grades in whatever formula they use. (I have alluded to this already.) But what is clear is that a student scoring very high on the SAT is more apt to be selected than individuals with lower scores, high school grades notwithstanding. Wing and Wallach, using only the SAT scores, were able to predict actual admissions decisions for several selective institutions at a 76 percent level.[7]

But what do the test scores tell us? They predict—and not very highly at that—first year grades. And their predictive value is not as high as that of high school grades. How important is that prediction? Does it relate as closely to the full baccalaureate program? And do the grades being predicted have much relationship to success in one's adult life?

Bowdoin conducted several studies which showed little relation between SAT scores and final standing and ultimately dropped the SAT as a requirement. To my knowledge, there has been no decline in quality at Bowdoin—though the population selected has broadened.[8] I don't believe that I need to provide for this group the extensive literature on the relationship of college grades to adult achievement.[9] Hoyt's reviews, I suspect, are now quite commonplace.

An even larger literature than was provided by Hoyt, however, exists in the medical field.[10] The entrance examination—similar in format to other admissions tests—has demonstrated little relation to medical school grades, medical school grades have little relationship to practice nor, for that matter, do scores on medical board examinations.

Are there other definitions of talent that might function as well as, possibly better than, grades and test scores as predictors of success in college as well as later life? Holland and Richards in a number of studies for ACT confirm that extracurricular, non-academic achievement/accomplishment can be identified as independent forms of talent, having negligible correlations to grades, aptitude or academic achievement tests.[11] They suggest that "if a college wishes to find students who will do outstanding things outside the classroom and in later life, then more attention needs to be given to non-academic accomplishment." How important is such a perspective? I would think enormously important. This is the major point of Wing

---

Footnotes continued from last page

Review, Nov. 1972, pp. 521-562 and "Social Class, Academic Ability and College 'Quality' ", Social Forces March 1975, pp 381-398. See also L. C. Solmon and P. Wachtel, "The Effects on Income of Type of College Attended," Sociology of Education, Winter, 1975 pp. 75-90.

[7] Cliff Wing and Michael Wallach, College Admissions and the Psychology of Talent, Holt Rinehart and Winston: N.Y., 1971. That Wing and Wallach were able to demonstrate so forceably the power of the SAT in actual admissions decisions doesn't support—in spite of the claims that are made—the perspective that it plays a "small, overestimated role."

[8] As the SAT declined in importance for purposes of selection at Harvard between 1958 and 1968, grades of students increased. 40 percent of the 1958 class was on the Dean's List compared to 51.8 percent in 1968. There was 10.6 percent failure among the class of 1958 compared to 4.8 percent for the class of 1968 (Whitla, p. 485). And it should be noted that a number of selective institutions never made use of the SAT or other instruments of that kind. For example, Bennington and Hampshire Colleges.

[9] Don Hoyt summarized 46 predictive studies concluding that "present evidence strongly suggests that college grades bear little or no relationship to any measure of adult achievement." The Relationship Between College Grades and Adult Achievement: A Review of the Literature, American College Testing Service, Report No. 7, Sept. 1965.

[10] John Wingard and John Williamson, "Grades as Predictors of Physicians Career Performance: An Evaluate Literature Review," Journal of Medical Education, Vol. 48, pp. 311-322.

[11] John Holland and James Richards, Academic and Non Academic Accomplishment in a Representative Sample Taken from a Population of 612,000, Research Report No. 12, American College Testing Program, May 1966 and Academic and Non Academic Accomplishment: Correlated or Uncorrelated? Research Report No. 2, ACT, April 1965. See also Leo Munday and J. C. Davis, Varieties of Accomplishment After College: Perspectives on the Meaning of Academic Talent, Research Report No. 62, ACT, March 1974.

and Wallach's research on College Admissions and the Psychology of Talent as well as that of David McClelland.[12] How one defines talent—the weight that is placed on one form of talent in relation to other forms of talent—has a potentially significant impact on our society, especially when important decisions result from that definition.

When Wing and Wallach applied to Duke's application pool a preference to non-academic accomplishment,[13] with minimum SAT scores set at the 50th percentile and with high school grades falling among the upper two-thirds of a class, they ended up with a selected population radically different—broader in every respect—than the population actually selected by a formula which weighted only SAT and high school grades. This suggests, I believe, a reasonable case for seeking in greater detail what have tended to be non-traditional data and to use that data in our academic selection procedures. That we have tended to place more emphasis on test scores and grades than seems justified is becoming more clear and the need to retreat from such an emphasis is gaining adherents.

Being an historian, I tend to look for earlier efforts which parallel in some ways current circumstances. As we think about alternatives to standardized testing formats, we might wish to refer again to the neglected eight year study which involved Wilford Aiken and Ralph Tyler in the 1930's.[14] Twenty-nine secondary schools were involved in developing alternative educational formats that did not relate to the traditional requirements for college entrance. These schools presented to colleges, on behalf of students who wished to go on to college, a portfolio outlining a broad range of an individual's accomplishments. They provided considerably more information than we see today and what was presented had a potential for use that current admission's data lack.[15]

What would students present to a college if they wished to demonstrate their fullest capabilities? [16] It might be useful to examine the implications of such a question. The College Board might enter into some experimentation with a number of selected schools, seeking descriptive formats for use within member institutions rather than the SAT or various achievement test batteries. We might in the process learn a good deal more than we now know about our entering students and thereby provide more appropriate advisement and placement.[17] Secondary schools might also begin to initiate alternative ways to present the strength of their students, using in the process a broad view of talent. By bringing much more experimentation to this system, we are certainly not going to increase the problems which higher education already experiences and we may find more appropriate mechanisms for doing what colleges and universities need to do better.

Being particularly conscious of time and recognizing that I have not addressed all of the concerns which were to make up this session, I will now comment on several other issues, if only to raise a number of questions.

We are witnessing in our colleges and universities a resurgence of interest among the 21-40 year old population. This interest manifests itself in terms of particular courses as well as degree programs. Are our typical testing formats appropriate for this population? Should these students be required to submit a score? To what end? Is their interest and motivation for the course or program sufficient? Do we have sufficient vehicles for them to demonstrate the quality of their experiential learning? CLEP—and the American Council on Education has just entered into a process of validation for other similar programs—is certainly a useful direction, one that has been used successfully by thousands of individuals, young and old. But is the format responsive to the learning of older persons who have been removed from educational settings for five or more years? I personally don't believe so.[18]

---

[12] Wing and Wallach give particular credence to what they call accomplishment characteristics—political leadership, athletic leadership, work experience, acting, writing, painting—behaviors which are ends in and of themselves. They conclude, as have others, that such factors act quite independently from grades and test scores. See also David McClelland, "Testing for Competence Rather than Intelligence," American Psychiatrist, Jan. 1973 and Michael Wallach, "Tests Tell Us Little About Talent," American Scientist, Jan/Feb 1976.

[13] Political leadership, athletic leadership, successful work experience, painting, acting, performing on a musical instrument, singing, etc.

[14] Wilford Aiken summerizes the effort in The Story of the Eight Year Study, McGraw-Hill: N.Y., 1942.

[15] I am aware that many schools have large numbers of applicants and because of limited admissions staff tend to desire simple quantification over descriptive portfolios. I understand this but don't accept it as responsible practice.

[16] I had that experience in relation to a selective experimental college. It was not a large enterprise but one that provided applicants, and their sponsors, room to elaborate, explain, present the broadest possible perspective. It was a process that proved particularly useful to the faculty and was viewed as a constructive experience for the students.

[17] College advising, especially at the freshmen year, tends now, in most settings, to be quite undifferentiated. In part this occurs because colleges do not devote enough of their resources to personal advisement. But another factor is that advisors have very little information upon which to differentiate. While some persist in believing that a high or low math score on the SAT suggests a particular math sequence for a student, such a belief may not contribute to the most appropriate advisement or placement.

Would we have more commitment to broader definitions of talent—to alternative means of presenting competences and accomplishments—if tests such as the SAT and ACT were more accessible, published along with their correct answer form 30 days after administrations as a recent California Senate Bill proposed?[19] Would students and counselors alike benefit from an item by item review? Again, I believe so. The single score which now comes back doesn't reveal very much about how a student dealt with the test. Is it possible that the student completed one-third of the testing and of that one-third responded "correctly" most of the time? Is it possible that a student scoring 600 and a student scoring 500 did similarly on all but a very small section of the test? Is it possible that a student provided insight into many of the questions that needs to be reviewed?[20]

If the tests were open, testing might become less mystical, less powerful in selection and possibly more useful fur purposes of placement advising. Would such an effort influence high school curriculum to an even higher degree than is now the case? That is a possibility if colleges and schools continue their lack of significant dialogue and lack of intensity in seeking more appropriate articulation.

Enough for now. As the typical college age population declines and as our economy changes, we need to address more freshly issues relating to college admissions, placement and advanced credit. It is a time to establish a range of alternative directions, for finding an even better balance between egalitarianism and meritocracy, for assuring access to the best resources of higher education for all who deserve and can benefit from such resources.[21] I am aware that I have highlighted the complexities and the more negative possibilities growing out of the tests. I did so purposefully lest we forget where we have been and how far we need to go. Having said that, I can also share with you my optimism for the future. I am convinced that there will be more modesty in the years ahead about what can be measured and that testing as we know it will moderate as a significant influence in our society.

Chairman PERKINS. I thank you very much for your testimony. We will now hear from Dr. Frank Snyder, representing the Association of American Publishers.

STATEMENT OF FRANK SNYDER, McGRAW-HILL, REPRESENTING ASSOCIATION OF AMERICAN PUBLISHERS, ACCOMPANIED BY ROY H. MILLENSON, AAP, WASHINGTON OFFICE

Dr. SNYDER. Thank you, Mr. Chairman.

My name is Frank Snyder.

I am the publisher of CTB/McGraw-Hill of Monterey, Calif., a commercial firm which develops and distributes a variety of testing instruments, and I am testifying here today on behalf of the Association of American Publishers.

---

[18] The College Board might wish to give attention to the efforts being made by institutions such as Thomas Edison in New Jersey, Empire State in New York and Metropolitan State in Minnesota which are seeking a variety of means to certify for college credit adult learning.

[19] The new SAT Student Booklet, unlike previous versions, prints an entire facsimile of a SAT. This is a step forward but it is still not the test and it cannot be used by the participants in a manner suggested by the California Bill.

[20] For a period, ETS suggested that these tests, being aptitude tests, were not particularly amenable to significant score shifts from preparation programs. But a large industry has grown up around the belief that 20-30 hours of instruction can make a significant difference; that scores can be improved by 100 points or so. These schools have become, over the past two years, the subject of a major FTC review in relation to "truth in advertising." While I do not yet know the content of the report, I suspect that the review will uphold the proprietary school's general claim. The College Board itself has recently published a report which supports that the concepts, principles, etc., which make up the math section of the SAT can be taught and students who participated in a 21 hour course designed for that purpose significantly improved their scores from pre-test to post-test.

[21] Many in our country believe that access to higher education is now open to all who can benefit. Humphrey Doerman's, Toward Equal Access (Bush Foundation and College Board, 1978) suggests otherwise. Doerman notes that some 200,000 qualified students per year are not in post high school schooling because financial assistance is not sufficiently available. While we have made giant strides in financial programs for young people who wish further education, we have not yet reached a level that is equal to our capacity.

I am accompanied by Roy H. Millenson of the AAP Washington office.

AAP, which is the general association of publishers in the United States, includes among its members companies which comprise the commercial testing community. Harcourt Brace, Jovanovich, which testified here last month, and Science Research Associations, however, companies which Representative Weiss mentioned in his opening statement, are not members of our association.

Our members are the source of tests which are not only used in elementary and secondary education, but also by private enterprise for hiring and promotion. At this time our members do not provide tests for college admittance, although we would not preclude the possibility of entry into that field at some future time.

While only H.R. 3564 applies specifically to tests produced by AAP constituent companies, our concerns must necessarily encompass H.R. 4949 inasmuch as some of its provisions might well be incorporated in any final legislation applying to job entry testing. Our testimony therefore deals with each of these measures in some detail.

Notwithstanding our agreement with the purposes which we infer to be implicit in the drafting of the proposed legislation, we doubt that there is a demonstrated need for the proposed legislation that is not currently fulfilled in principle and recommended practice in the form of standards for educational and psychological tests produced by professional organizations such as the American Psychological Association, the American Educational Research Association, and the National Council on Measurement in Education, as well as in currently available information provided by testing agencies, as an ongoing attempt to provide information to examinees.

GIBBONS BILL (H.R. 3564)

Now, we have prepared a rather detailed list of analyses of the bill section by section. I would like to provide some summary comments, first on the Gibbons bill.

There are three main things that concern us about the Gibbons bill.

The first one is the broadness of the bill. Section 2(3) includes oral examinations within the definition of "test." This could present difficulties both in enforcement and definition. For example, is a person interviewed by a personnel director or a college admissions officer being subjected to a "test"?

Our contention under the proposal contained within the bill that this would lead to a very complicated situation.

In addition to the broadness of the definition of test contained within the bill, the bills are also designed to apply to a very broad variety of testing programs; namely, admissions programs, occupational testing programs, and proprietary testing programs conducted by individual companies due to their participation in interstate commerce.

We see that the need has not been demonstrated for regulation in these areas of testing, and we are very concerned about the possible unanticipated consequences of such legislation applied to such programs.

Also we would point out to the committee that the compliance provisions of the bill are not clear.

Admissions procedures to institutions of higher learning and to occupations are represented, in practice, by a variety of arrangements between agencies which produce tests in institutions, organizations, and commercial concerns which use these tests.

In the area of occupational admissions, the test producer usually is not directly or indirectly involved in the actual administration of the examination, and thus would not be in a position to comply with all the disclosure provisions in this section.

We would also like to point out to the committee that if our interpretation of the bill is correct, in section 2(5) we would raise the question of whether or not society includes labor unions, and does this bill intend to govern tests for admittance into labor unions, when such admittance controls entrance into an occupation.

Our second concern is a presentation of results.

The bill has several stipulations in it which we contend are not in agreement with currently accepted practice within the professional community; or that would lead to unanticipated consequences that would make interpretation more rather than less difficult.

For instance, ranking of students in relation to total test score may not fit in with the logic upon which that particular test was constructed and would tend to be in conflict with currently accepted procedures of considering test scores as elements in an overall profile characterizing an examinee rather than a single determinance for admittance.

Section 6(b)(3) requires that the testing agency specify cutoff scores for admission to institutions of higher education or occupations. This is unreasonable both from a technical viewpoint and in terms of a testing agency's capability to fulfill the requirement. Such cutoffs are generally unknown to the testing agency in both academic admissions and occupational testing programs.

Decisions on cutoff scores are under the control of the admitting or hiring institution and would logically change with changing levels of achievement or aptitude, size of the examinee population, and available jobs or space.

The use of single cutoff scores is discouraged by the testing agencies based on their potential misuse as a mechanistic way of using test information.

The use of cutoff scores would also tend to minimize the use of other criteria and would be detrimental to considerations placed on criteria other than test results contained in the totality of a potential entrant's abilities.

When applied in the occupational testing area the use of cutoff scores would also tend to lead to a situation in which no differential decisions, based on test performance, could be made above the cutoff score. In the long run this situation could eventuate in random hiring of minimally qualified individuals rather than the hiring of those deemed best qualified.

I would like to add one more thing to that. There is also a stipulation in the bill that applies a differential treatment of scores for ability tests and for achievement tests.

Basically the stipulation is that no results from achievement tests can be scored in such a manner that the distribution of the respondent's performance is used.

This differential treatment implies a distinction in theory between the way the two types of tests are treated that is not in agreement with current professional thinking in testing methodology.

The requirement that no test which measures knowledge or achievement shall be graded on the basis of the distribution of scores of other test subjects also assumes a static level of ability in our population over time, and/or a static capacity across institutions to accept admissions or hire applicants. Both assumptions are neither desirable nor supportable.

Further, if such a requirement were promulgated in occupational testing, it would assume the existence of a definitive list of minimum skills necessary for job entry.

Such definitive lists do not currently exist and their absence would, we believe, lead to pressures at the Federal level to mandate what such a list would include; or, in effect, a Federal specification of minimum job entry skills. This would clearly be a violation of both institutional and State prerogatives and should be avoided.

If this bill is passed into law we recommend that the earlier date be no earlier than 12 months after enactment to provide adequate time for the promulgation of the final regulations and subsequent alteration of test materials in order to comply with the law.

In conclusion, we feel enactment of H.R. 3564 would result in myriad foreseen and unforeseen consequences and would present dilemmas of enforcement and compliance.

WEISS BILL (H.R. 4949)

Our comments on H.R. 4949 necessarily relate only to those provisions that might apply to occupational admission testing, should the committee decide to combine H.R. 4949 and H.R. 3564 as has been suggested.

Specifically, the bill mandates that at least three different types of information not commonly available now be provided. They first require in section 3(a)(2) that the correlation between a test score and success in a career be provided to examinees.

Beyond the realization that the presentation of correlations to test subjects is unlikely to communicate information that the test subject can understand, the committee should note that meaningful correlations between test scores and career success are almost never available due to the difficulty of defining what is meant by career success.

The only way I could envision an adequate implementation of this provision is if professional people were tested year after year and those results correlated back to the entry level test. While this might be a reasonable thing it would certainly be beyond what has currently been achievable in practice.

Section 3(a)(4)(B) of the bill also requires the comparison of test scores by income group. Information necessary for this analysis, if not routinely available, would require much more paperwork, would potentially be viewed as an invasion of privacy, and would probably generate more misunderstandings about the tests.

It is in fact true that test scores do tend to correlate with income groups. Given the way our society is structured that kind of conclusion should not be a surprise to anybody.

Section 3(a)(4)(C) requires the expression as a percentage of the improvement in test scores as a result of a test preparation course. The information to form such a statistic is not likely to be available to the test producer sufficient for any meaningful interpretation to be made and would depend almost entirely on the contents of a specific test preparation course when compared to a specific test.

In a sense the statistic itself would be almost meaningless; since it would depend almost entirely on the contents of a specific course, a specific test and a particular institution within which the correlation was generated.

Section 3(a)(9) refers to handicapped test subjects, a term which is not defined. The committee will recall the difficulties that ensued because of the lack of a clear definition of "handicapped" in section 504 of the Vocational Rehabilitation Act of 1973—Public Law 93-122. That 43-word section eventually resulted in some 26 pages of regulations.

A second concern is the release that is being demanded of what a proprietary testing company such as my own must consider to be competitive data.

Section 4(a)(1)(B) indicates that we have to release the terms of our contracts of customers. We can only consider this an invasion of privacy—it is violative of the spirit, if not the letter, of the Privacy Act of 1974, Public Law 93-579—and one that would place any company complying with this provision at a distinct competitive disadvantage.

This provision is not in the New York State statute or in any of the testing bills proposed in the States, and does not belong in any Federal statute.

Section 7(2)(F) would oblige testing agencies to reveal their revenue from each testing program. This is proprietary information which would only serve the purpose of informing competitive companies and would place the reporting company at a competitive disadvantage. This provision is not in the New York statute or in any of the bills proposed in the States and also does not belong in any Federal law.

A final topic that I would like to comment on is the disclosure of test questions. We see two problems. I am not commenting on college admissions testing. First, the security of the test for the purchasing company's purpose would be breached.

Proprietary testing in the United States is a somewhat different animal than admissions testing to institutions of higher education. The tests are frequently used by individual companies to determine eligibility for employment by the company itself. They are administered frequently, usually many times a year and usually to a group of no more than several dozen applicants.

As a matter of fact, those tests sold by my company within this category rarely exceed a quantity of 1,500 copies to an individual purchaser for use over a period of time.

The security of the test if the items are required to be released for the purchasing company's purposes would be breached. It would be infeasible and uneconomical and considering the quantities, a great bother to our customer to substitute a new test to begin within a month or so and to repeat that procedure again and again throughout the year.

Beyond financial and administrative considerations it is also highly probable that no replacement tests would be available because of the costs. The data base which we must as a professional community provide information on would increase to the point where it would not be financially viable to continue the publication of occupational data.

Second, the tests in question are copyrighted. The entire issue of test copyrights should be explored by the committee. We strongly oppose any application of section 5 to proprietary occupational entry tests.

In section 7(2) with respect to paragraphs (A), (B), (C), and (D), this information is not routinely available to proprietary testing agencies. These provisions are not in the New York statute.

One further point, if H.R. 4949 is enacted, the effective date should be no earlier than 12 months after enactment to provide for promulgation of the final regulations and subsequent alteration of test materials to comply with the law.

Since H.R. 4949 contains major provisions which are not in the New York statute, it cannot be argued that as a result of complying with the New York law testing agencies will have no need for additional time to prepare for compliance with a Federal statute.

FEDERAL DOUBLE STANDARD

I would like to comment now on the Federal role in testing which we see as the imposition of a double standard. The committee should consider the relationship between the legislation under discussion and current practices of the Federal Government with respect to testing. According to the Office of Personnel Management, in fiscal year 1978, 1,616,178 persons took Federal civil service tests. The pending legislation requires that job entry examinations meet certain prescribed requirements.

It seems strange that the Congress should impose these requirments upon the private sector—requirements which themselves have not undergone trial—without first imposing them upon tests given by the Federal Government itself.

If the testing practices to be imposed by the pending legislation are, in fact, important to the rights of individuals and the national interests, it seems only proper that the first step should be taken by the Federal Government itself in the interests of the more than 1.6 million individuals taking civil service examinations each year.

Beyond civil service examinations, it should also be noted that the armed services have extensive examination procedures both at the point of entry and for subsequent promotion decisions.

That Federal occupational entry examinations have their shortcomings is pointed up by the May 15, 1979, report to the Congress by the Comptroller General.

Also, we would like to invite the committee's attention to section 3 of the Privacy Act of 1974 (Public Law 93-579) which added a new section 522a to title V of the U.S. Code. Subsection (k)—"Specific Exemptions"—of that section permits an agency head to "exempt any system of records within that agency" from certain disclosure and access requirements of the statute. The law includes among these exemptions the following:

> (6) testing or examination materials used solely to determine individual qualifications for appointment or promotion in the Federal service, the disclosure of which would compromise the objectivity or fairness of the testing or examination process;

This provision, to put it plainly, points up the fact that the proposed legislation would require the private sector to do something which Federal agencies are not required to do under similar circumstances.

RECOMMENDATIONS

The enactment at this time of H.R. 3564 and the application of H.R. 4949 to occupational entry testing is premature. It is clear that the entire issue needs further study. For example, such a key issue as the copyright of test questions has little been explored by the committee.

In addition, since the New York statute goes beyond the California law and the proposed Federal statute goes beyond the New York law and proposals pending in the States, it would be wise to garner some experience with the State laws before proceeding at the Federal level.

There is also a study underway by the National Academy of Sciences on ability testing which reports should be out within a year. It would seem prudent to examine that committee's findings before any legislation is enacted governing testing.

The second recommendation: The requirements for testing agencies contemplated by the pending legislation should first be mandated for the standardized tests administered by the U.S. Office of Personnel Management, formerly the Civil Service Commission, the armed services, and other Federal agencies.

If it is a matter of public interest to apply certain testing standards to the private sector, it is even more in the public interest that the Federal Government should adhere to those standards.

We should like to reiterate one legislative recommendation in the area of testing which we made in testimony submitted to the Postsecondary Education Committee of the House Committee on Education and Labor. We urge that section 532(a)(2)(B) of the Higher Education Act be amended by adding "testing" after "development".

This amendment would authorize teacher centers to provide training to familiarize teachers with developments in testing as well as in curriculum development and educational research. It is complementary to part B, educational proficiency standards of ESEA title IX, as amended by the recently enacted Educational Amendments of 1978, and should further the understanding of the development and appropriate use of tests.

We ask that our brief testimony submitted to the Postsecondary Education Subcommittee be included in the record as part of this statement.

Our final recommendation: Section 922 of the Elementary and Secondary Education Act, as amended by the Education Amendments of 1978 (Public Law 95-561), provides for "training of and assistance to administrators, teachers, and other instructional personnel in the use of tests and test results" in the case of certain tests in the basic skills in elementary and secondary schools.

We recommend that consideration be given to providing assistance in the use of standardized tests and tests results at the post-secondary level and with respect to occupational entry examination used outside of schools.

CONCLUSION

The pending legislation is not, as a number of witnesses have implied, an extension to the Nation of the New York State statute. The New York statute does not cover occupational admissions tests, as does H.R. 3564. In H.R. 4949 there are both major and minor provisions which are not a part of, or which are significantly different from, the New York statute, including the following: (1) The findings in section 2, H.R. 4949; (2) Section 4(a)(1)(B); (3) Section 4(b); (4) Section 7 on testing costs and fees to students; (5) Section 8(b), the penalty provisions.

As for those who wish to abolish all standardized testing and might wish to expand the pending legislation to achieve that end, we are reminded that Irwin Polishook, vice president of the American Federation of Teachers, asserted that while the testing process has some flaws, abolishing standardized testing would be an emotional and damaging strategy that would deprive us of a valuable education tool, to which there is no alternative.

We are also impressed with the arguments questioning the constitutionality of the pending legislation and believe that this merits further exploration in sufficient detail.

Finally, it is evident from the testimony presented at these hearings that there is deep concern about the adequacy of examinations required for entrance to institutions of higher learning and the nonprofit organizations that produce them.

Regardless of the committee's findings in such areas, we are concerned that inappropriate actions will be taken against the proprietary testing community whose operations are of a much lesser magnitude, involve a totally different constituency, and whose trade practices are by nature markedly different.

[The prepared testimony of Frank Snyder follows:]

TESTIMONY OF DR. FRANK SNYDER,[1] ASSOCIATION OF AMERICAN PUBLISHERS

My name is Frank Snyder. I am the Publisher of CTB/McGraw-Hill of Monterey, California, a commercial firm which develops and distributes a variety of testing instruments, and am testifying here today on behalf of the Association of American Publishers. AAP, which is the general association of publishers in the United States, includes among its members companies which comprise the commercial testing community. Harcourt, Brace, Jovanovich (which testified here last month) and Science Research Associates, however, companies which Rep. Weiss mentioned in his opening statement, are not members of our association.

Our members are the source of tests which are not only used in elementary and secondary schools, but also by private enterprise for hiring and promotion. At this time our members do not provide tests for college admittance, although we would not preclude the possibility of entry into that field at some future time.

---

[1] Accompanied by Roy H. Millenson of the AAP Washington office.

While only H.R. 3564 applies specifically to tests produced by AAP constituent companies, our concern must necessarily encompass H.R. 4949 inasmuch as some of its provisions might well be incorporated in any final legislation applying to job entry testing. Our testimony therefore deals with each of these measures in some detail.

Notwithstanding our agreement with the purposes which we infer to be implicit in the drafting of the proposed legislation, we doubt that there is a demonstrated need for the proposed legislation that is not currently fulfilled in principle and recommended practice in the form of Standards for Educational and Psychological Tests by professional organizations such as the American Psychological Association, the American Educational Research Association, and the National Council on Measurement in Education, as well as in the form of currently available information provided by testing agencies, as an ongoing attempt to provide information to examinees.

First, we should like to comment on specific provisions of the pending bills, as follows:

H.R. 3564 (GIBBONS BILL)

1. *Sec. 2(3)* includes oral examinations within the definition of "test". This could present difficulties both in enforcement and definition. For example, is a person interviewed by a personnel director or a college admissions officer being subjected to a "test"?
2. *Sec. 2(4)* defines "test score" as a numerical value. Besides the obvious complications with respect to oral examinations, is it the intent of this legislation to mandate that all procedures be reduced to numerical values? This would seem to preclude a simple "pass-fail" or other non-numerical score. On the face of it, this provision would seem to be unwarranted given the current statements of need, and to have the potential for unanticipated consequences at variance with the central purpose of providing information and insuring equitable treatment.
3. *Sec. 2(5)* includes in the definition of "persons" the term "societies". Does this include labor unions? Does this bill govern tests for admittance into labor unions, especially when such admittance controls entrance into an occupation?
4. *Sec. 6* raises questions as to who specifically shall be required to furnish the information required in subsections (b) and (c). Admissions procedures to institutions of higher learning and to occupations are represented, in practice, by a variety of arrangements between agencies which produce tests and the institutions, organizations and commercial concerns who use these tests. In the area of occupational admissions, the test producer usually is not directly or indirectly involved in the actual administration of an examination, and thus would not be in a position to comply with all the disclosure provisions of this section.
5. *Sec. 6(a)(1) and (2).* It is not clear what is meant by a "detailed" description.
6. *Sec. 6(a)(3).* The term "reliability" has not been defined. Given the number of interpretations of this concept as it currently applies to tests, it should be more thoroughly examined. Further, the Committee should note that when an examination is prepared for a particular purpose, reliability information may not be available prior to the administration of the examination.
7. *Sec. 6(b)(2).* The requirement that individuals be "ranked" in relation to other individuals is inconsistent with Section 6(c) which prohibits grading of tests of knowledge or achievement on the basis of the score distribution of examinees. Ranking individuals by total test performance may also be contrary to the logic on which the test was constructed and in conflict with currently accepted procedures of considering test scores as elements in the overall profile characterizing an examinee.
8. *Sec. 6(b)(3)* requires that the testing agency specify cut-off scores for admission to institutions of higher education or occupations. This is unreasonable both from a technical viewpoint and in terms of a testing agency's capability to fulfill the requirement. Such cut-offs are generally unknown to the testing agency in both academic admissions and occupational testing programs. Decisions on cut-off scores are under the control of the admitting or hiring institution and would logically change with changing levels of achievement or aptitude, size of the examinee population, and available jobs or space. The use of single cut-off scores is discouraged by the testing agencies based on their potential misuse as a mechanistic way of using test information. The use of cut-off scores would also tend to minimize the use of other criteria and would be detrimental to considerations placed on criteria other than test results contained in the totality of a potential entrant's abilities.

When applied in the occupational testing area, the use of cut-off scores would also tend to lead to a situation in which no differential decisions, based on test performance, could be made above the cut-off score. In the long run this situation could eventuate in random hiring of minimally qualified individuals rather than the hiring of those deemed best qualified.

9. *Sec. 6(a)(4)* should be deleted because the term "further information" is unclear.

10. *Sec. 6(c).* In addition to the conflict between this section and Section 6(b)(2), this section also calls for the differential treatment of the scores obtained from achievement tests and aptitude tests. Such differential treatment implies a distinction in theory between the way the two types of test are treated that is not in agreement with current professional thinking in testing methodology. The requirement that no test which measures knowledge or achievement shall be graded on the basis of the distribution of scores of other test subjects also assumes a static level of ability in our population over time, and/or a static capacity across institutions to accept admissions or hire applicants. Both assumptions are neither desirable nor supportable. Further, if such a requirement were promulgated in occupational testing, it would assume the existence of a definitive list of minimum skills necessary for job entry. Such definitive lists do not currently exist and their absence would, we believe, lead to pressures at the federal level to mandate what such a list would include; or, in effect, a federal specification of minimum job entry skills. This would clearly be a violation of both institutional and state perogatives and should be avoided.

11. If this bill is passed, in its present form or in an altered form, the effective date of the act should be no earlier than twelve months after enactment to provide adequate time (a) for the promulgation of the final regulations and (b) subsequent alteration of test materials in order to comply with the law.

In conclusion, we feel enactment of HR 3564 would result in a myriad of foreseen and unforeseen consequences and would present dilemmas of enforcement and compliance.

H.R. 4949 (WEISS BILL)

Our comments on HR 4949 necessarily relate *only* to those provisions that might apply to occupational admission testing, should the Committee decide to combine HR 4949 and HR 3564 as has been suggested. Our specific comments follow:

1. *Sec. 3(a)(3)* calls for providing the test subject with the correlations between test scores and "success" in a career. Beyond the realization that the presentation of correlations to test subjects is unlikely to communicate information that the test subject can understand, the Committee should note that meaningful correlations between test scores and career success are almost never available due to the difficulty of defining what is meant by "career success." This provision, if implemented for occupational testing, would mandate determining "success" in a carrer for which admittance is sought. Would this require that each profession or occupation be tested or reviewed over the years throughout the career span in some measurable from so a correlation would be generated between that score and an individual's test score? This provision is unclear and unenforceable in the present context of occupational admissions testing.

2. *Sec. 3(a)(4)(B)* requires the comarison of test scores by income group. Information necessary for this analysis if not routinely available would require much more paper work, would potentially be viewed as an invasion of privacy, and would probably generate more misunderstandings about the tests.

3. *Sec. 3(a)(4)(C)* requires the expression as a percentage of the improvement in test scores as a result of a test preparation course. The information to form such a statistic is not likely to be available to the test producer sufficient for any meaningful interpretation to be made and would depend almost entirely on the contents of a specific test preparation course when compared to a specific test.

4. *Sec. 3(a)(9)* refers to "handicapped test subjects", a term which is not defined. The Committee will recall the difficulties that ensued because of the lack of a clear definition of "handicapped" in Sec. 504 of the vocational Rehabilitation Act of 1973 (PL 93-122). That 43-word section eventually resulted in some 26 pages of regulations.

5. *Sec. 4(a)(1)(B).* We can see no reason why any proprietary testing company should be obligated to reveal to its competitors the terms of contracts with its customers. This is an invasion of privacy[2] and one that would place any company complying with this provision at a distinct competitive disadvantage. This provision is not in the New York State statute or in any of the testing bills proposed in the states, and does not belong in any Federal statute.

5. *Sec. 5* presents particular problems for proprietary test publishers. These tests, when used for admittance into an occupation, are frequently administered by individual companies to determine eligibility for employment by the company itself.

---

[2] It is violative of the spirit, if not the letter, of the Privacy Act of 1974 (Public Law 93-579).

They are administered frequently, usually many times during a year, and usually to a group of no more than several dozen applicants. As a matter of fact, these tests are by my company in quantities that rarely exceed 1,500 copies to an individual purchaser for use over a period of time.

We see two problems, if disclosure of test questions and answers are mandated in the occupational testing area. First, the security of the test, for the purchasing company's purposes, would be breached. It would be unfeasible and uneconomical—and, considering the quantities, a great bother to our customers—to substitute a new test to be given within a month or so and to repeat that procedure again and again throughout the year. Beyond financial and administrative considerations, it is also highly probably that no such repalcement tests would be available because the cost of test development and data gathering to support reliability and validity would increase to the point where it would not be financially viable to continue the publication of occupational admission tests. Secondly, the tests in question are copyrighted. HR 4949, if applied to occupational entry testing, would force the test publisher to, in effect, give away his copyrighted material substantially before the expiration of its useful life. The entire issue of test copyright should be thoroughly explored by the Committee.

We strongly oppose any applications of Section 5 to proprietary occupational entry tests.

7. *Sec. 7(2).* With respect to paragraph (A), (B), (C), and (D), this information is not available to proprietary testing agencies. It should also be noted that these mandatory provisons are not in the New York State law.

8. *Sec. 7(2)(F)* would oblige testing agencies to reveal their revenue from each testing program. This is proprietary information which would only serve the purpose of informing competitive companies and would place the reporting company at a competitive disadvantage.[3] This provision is not in the New York statute or in any of the bills proposed in the states and also does not belong in any Federal law.

9. *Sec. 7(2)(G)* also mandates the release of proprietary information which would only be to the advantage of the competitors of the company providing it and which would place it at a competitive disadvantage. This provision is not in the New York statute or in any of the bills introduced in the states and also does not belong in any Federal law.

10. *Sec. 8(a)* is inconsistent with Section 431 of the General Education Provisions Act. Section 431 provides that within sixty days after the enactment of legislation, the Commissioner of Education must submit to Congress a schedule showing when it is planned to promulgate final regulations. These final regulations must be issued within 180 days after the submission of the schedule. It should be noted that the Office of Education has difficulty complying with this provision. Final regulations have yet to be issued for parts of the Education Amendments of 1979 (Pub. L. 95-561) signed into law on November 1, 1978.

11. If H.R. 4949 is enacted, the effective date should be no earlier than twelve months after enactment to provide for promulgation of the final regulations and subsequent alteration of test materials to comply with the law. Since H.R. 4949 contains major provisions which are not in the New York statute, it cannot be argued that as a result of complying with the New York law testing agencies will have no need for additional time to prepare for compliance with a Federal statute.

THE FEDERAL ROLE IN TESTING—A DOUBLE STANDARD

The Committee should consider the regulationship between the legislation under discussion and current practices of the Federal Government with respect to testing. According to the Office of Personnel Management, in fiscal year 1978 1,616,178 persons took Federal civil service tests. The pending legislation requires that job entry examinations meet certain prescribed requirements. It seems strange that the Congress should impose these requirements upon the private sector—requirements which themselves have not undergone trial without first imposing them upon tests given by the Federal government itself. If the testing practices to be imposed by the pending legislation are, in fact, important to "the rights of individuals and the national interests", it seems only proper that the first step should be taken by the Federal government itself in the interests of the more than 1.6 million individuals taking civil service examinations each year. Beyond civil service examinations, it should also be noted that the armed services have extensive examination procedures both at the point of entry and for subsequent promotion decisions.

That Federal occupational entry examinations have their shortcomings is pointed up by the May 15, 1979 report to the Congress by the Comptroller General, "Federal

---

[3] It is violative of the spirit, if not the letter, of the Privacy Act of 1974 (Public Law 93-579).

Employment Examinations: Do They Achieve Equal Opportunity and Merit Principle Goals?" (FPCP-79-46).

Also, we would like to invite the Committee's attention to Section 3 of the Privacy Act of 1974 (Pub. L. 93-579) which added a new Section 522a to Title V of the U.S. Code. Subsection (k)—"Specific Exemptions"—of that section permits an agency head to "exempt any system of records within that agency" from certain disclosure and access requirements of the statute. The law includes among these exemptions the following:

(6) testing or examination materials used solely to determine individual qualifications for appointment or promotion in the Federal service, the disclosure of which would compromise the objectivity or fairness of the testing or examination process;

This provision, to put it plainly, points up the fact that the proposed legislation would require the private sector to do something which Federal agencies are not required to do under similar circumstances.

RECOMMENDATIONS

1. The enactment at this time of H.R. 3564 and the application of H.R. 4949 to occupational entry testing is premature. It is clear that the entire issue needs further study. For example, such a key issue as the copyright of test questions has little been explored by the Committee. In addition, since the New York statute goes beyond the California law and the proposed Federal statute goes beyond the New York law and proposals pending in the states, it would be wise to garner some experience with the state laws before proceeding at the Federal level.

2. The requirements for testing agencies contemplated by the pending legislation should first be mandated for the standardized tests administered by the U.S. Office of Personnel Management (formerly the Civil Service Commission), the armed services, and other Federal agencies. If it is a matter of public interest to apply certain testing standards to the private sector, it is even more in the public interest that the Federal Government should adhere to those standards. In addition, since this is untried legislation, going considerably beyond current state statutes, application to the Federal establishment would serve as a proving ground before application is extended.

3. We should like to reinterate one legislative recommendation in the area of testing which we made in testimony submitted to the Postsecondary Education Committee of the House Committee on Education and Labor. We urge that Section 532(a)(2)(B) of the Higher Education Act be amended by adding "testing" after "development". This amendment would authorize teacher centers to provide training to familiarize teachers with developments in testing as well as in curriculum development and educational research. It is complementary to Part B (Educational Proficiency Standards) of ESEA Title IX, as amended by the recently enacted Educational Amendments of 1978, and should further the understanding of the development and appropriate use of tests. We ask that our brief testimony submitted to the Postsecondary Education Subcommittee be included in the record as part of this statement.

4. Section 922 of the Elementary and Secondary Education Act, as amended by the Education Amendments of 1978 (Pub. L. 95-961), provides for "training of and assistance to administrators, teachers, and other instructional personnel in the use of tests and test results" in the case of certain tests in the basic skill in elementary and secondary schools. We recommend that consideration be given to providing assistance in the use of standardized tests and test results at the post secondary level and with respect to occupational entry examination used outside of schools.

CONCLUSION

The pending legislation is not, as a number of witnesses have implied, an extension to the nation of the New York State statute. The New York statute does not cover occupational admissions tests, as does HR 3564. In HR 4949 there are both major and minor provisions which are not a part of, or which are significantly different from the New York statute, including the following:

1. The findings in Section 2, HR 4949;
2. Section 4(a)(1)(B), reference to which is made in this testimony;
3. Section 4(b), reference to which is made in this testimony;
4. Section 7, (Testing Costs and Fees to Students); and
5. Section 8(b), the penalty provision.

It should also be noted that the provisions cited above are not, for the most part, in any of the parallel bills which have been introduced in the states.

As for those who wish to abolish all standardized testing[4] and might wish to expand the pending legislation to achieve that end, we are reminded that Irwin Polishook, Vice President of the American Federation of Teachers, asserted that while the testing process has some flaws, abolishing standardized testing would be "an emotional and damaging strategy" that would "deprive us of a valuable education tool, to which there is no alternative."

We are also impressed with the arguments questioning the constitutionality of the pending legislation and believe that this merits further exploration in sufficient detail.

Finally, it is evident from the testimony presented at these hearings that there is deep concern about the adequacy of examinations required for entrance to institutions of higher learning and the nonprofit organizations that produce them. Regardless of the Committee's findings in such areas, we are concerned that inappropriate actions will be taken against the proprietary testing community whose operations are of a much lesser magnitude, involve a totally different constituency, and whose trade practices are by nature markedly different.

STATEMENT OF THE ASSOCIATION OF AMERICAN PUBLISHERS

The Association of American Publishers (AAP) is the general association of book publishers in the United States. It comprises the General Publishing Division, Direct Marketing/Book Club Division, Technical, Scientific and Medical Division, International Division, College Division, Mass Paperback Division, and School Division. Our 330 member publishing houses produce the vast majority of general trade, educational, reference, professional and religious books published in this country. AAP members publish 80 percent of the instructional materials used in the nation's classrooms.

SUGGESTED AMENDMENT

We urge that Section 532(a)(2)(B) of the Higher Education Act be amended by adding ", testing," after "development". A Ramyseur, showing how this amendment would alter existing law, is appended as Attachment A.

This amendment would authorize teacher centers to provide training to familiarize teachers with developments in testing as well as in curriculum development and educational research. It is complementary to Part B (Educational Proficiency Standards) of Title IX of the recently-enacted Education Amendments of 1978 (PL 95-561). The fact that an increasing number of States—36 at last count—require minimum competency testing makes this amendment highly desirable.

RATIONALE

Experience and the recent legislative mandates on school assessment indicate that testing is continuing to increase and is likely to increase even further in the next decade. Despite the criticisms and expressions of concern, it is apparent there are forces in education and society that are responding to pressures for better accountability and performance in the schools by requiring test data to determine the outcome of educational programs.

The latest report issued by the Education Commission of the States shows that thirty-six States have legislated the so-called minimum competency programs into being. Coupled with this are other on-going State and local assessment programs which indicate that testing continues to be a vital part of education.

A number of valid reasons why this addition would be both proper and beneficial relate to the general observation that the level of understanding and competence of many test users, whether teachers, counselors or administrators, leaves much to be desired. Certainly pre-service training in tests and measurements is seriously inadequate compared with, say, thirty years ago. Test publishers' efforts to deal with the problems through workshops, conferences, pamphlets and similar methods touch only a fraction of those who need it.

To look for a moment at the benefits for teachers:

With the increasing number of States requiring minimum competency testing, an understanding of the basic concepts of test construction in relation to the kinds of assessment outcomes desired is essential. If teachers are going to be expected to use the test results to identify, teach, and counsel students so they may achieve district-prescribed standards, then it seems only reasonable that the more knowledge teachers have about the testing process and the test results, the more meaningful will be the interface with their students.

---

[4] National Education Association Resolution 77-67 states: "The National Education Association strongly encourages the elimination of group standardized intelligence, aptitude, and achievement tests."

Some States are providing that proficiency assessments relate directly to locally-established standards and indicate that locally-developed criterion-referenced tests are the best way to insure that tests cover those skills and standards identified. When teachers are involved in the development of such tests, a good basic understanding of test construction is essential. Too often we find teachers thrust into such situations without sufficient training and then blamed for inadequate tests laced with ambiguous items, poor distractors, and personal bias.

Moving away from minimum competency testing, teachers, through the use of standardized test batteries, can determine areas wherein learning has been the most or least efficient. The analysis of learning rates and levels of accomplishment of students permits teachers to prescribe appropriate instructional programs.

By obtaining knowledge in statistics, even if just an overview, teachers will be able to summarize, describe or compare the results of measurement. This information provides more opportunities for accurate analysis of test results than simply assessing a single student's grade.

In the classroom one of the tasks of teachers is the quiz or teacher-made test. For a reliable and valid classroom test teachers need to know the basic concepts in order to properly construct and assess the adequacy of a given test.

Teachers can give a more informed voice to their school's administration if they have some knowledge of tests and measurements and can evaluate them as to whether the results will be as useful as predicted. in addition, teachers can more knowledgeably examine and recommend measures to be used throughout the school.

When teachers review student records, or those of transferees, knowledge of those tests for which scores are reported will lead to a better understanding of the student and his level of accomplishment.

When serving on a test selection committee, teachers with a good background knowledge of the wide variety of tests available, their application and how to best assess their usefulness make a valuable contribution to the school district and to the community.

In conclusion, for the above reasons and in the interest of improved understanding of educational measuement, we urge that Section 532(a)(2)(B) be amended to include testing.

APPENDIX A

TEACHER CENTERS

SEC. 532. (a)(1) The Commissioner is authorized to make grants to local educational agencies in accordance with the provisions of this section to assist agencies in planning, establishing, and operating tacher centers.

(2) For the purpose of this section, the term "teacher center" means any site operated by a local educational agency (or a combination of such agencies) which serve teachers, from public and non-public schools of a State, or an area or community with a State, in which teachers, with any assistance of such consultants and experts as may be necessary, may—

(A) develop and produce curricula designed to meet the educational needs of the persons in the community, area, or State being served, including the use of educational research findings or new or improved methods, practices, and techniques in the development of such curricula; and

(B) provide training to improve the skills of teachers to enable such teachers to meet better the special eductional needs of persons such teachers serve, and to familiarize such teachers with developments in curriculum development, testing, and educational research, including the manner in which the research can be used to improve their teaching skills.

Chairman PERKINS. Thank you very much.

We will now hear from Dr. Paul Pottinger, executive director of the National Center for the Study of Professions.

Go ahead, Doctor.

STATEMENT OF PAUL S. POTTINGER, EXECUTIVE DIRECTOR, NATIONAL CENTER FOR THE STUDY OF PROFESSIONS

Dr. POTTINGER. Thank you, Mr. Chairman, and members of the subcommittee.

I am pleased to have been invited to give my testimony today.

My name is Paul S. Pottinger. I am the executive director of the National Center for the Study of Professions here in Washington. I am testifying before you today in my official capacity as the center's director and spokesperson for Truth in Testing.

The center was established nearly 2 years ago as a private not-for-profit organization. We were started with a grant from the M. A. Rooney Foundation in Illinois. Currently, our funds come primarily from the National Institute of Mental Health for testing and manpower research in the mental health profession.

Our mission is partly to raise the level of awareness of the public about issues of manpower regulation within the professions. We study certification and licensing practices among other regulatory mechanisms.

We view tests as a major regulatory device in the professions, and we are concerned about the effects tests have on the freedom of competent citizens to pursue work of their choosing.

Part of the center's mission is also to help examine alternatives to currently used testing techniques. My research background is related to this. I have a doctorate in psychology from the University of North Carolina in Chapel Hill and a bachelor of science degree from Denison University in Ohio.

I was Senior Research Associate in the Office of Research at the National Institute of Education (DHEW) from 1973-74. For 3 years prior to establishing the National Center for the Study of Professions I was the director of assessment systems at McBer & Co., a Boston social science research and development consulting firm.

I recently edited a book titled "Defining and Measuring Competence"—Jossey-Bass, 1979. This book, including a chapter that I wrote, articulates some of the problems of current testing practices and the need for more and better information about tests.

So I come here today as both an advocate of proper test development and usage and as an expert. I am confident that better alternatives to our current testing practices will become possible if we can learn more about the tests at issue in these bills.

I have written testimony that in part, as I listened to my colleagues, I feel is redundant. I will try to reduce my comments to make points that have not already been made, if possible.

I would like to make the argument that testing is an essential, a critical gatekeeping mechanism in allocating opportunities for education and work, and because it is so critical we need the best information possible about what these tests measure, how valid they are, what effects they are having on the public, who are affected by taking them.

I think the evidence is fairly convincing that the test companies simply do not want this information known because the knowledge of these tests and their failings I think is threatening to their credibility as well as the revenue that they depend on for selling.

I would like to make some comments about multiple choice tests in general.

We are using a very poorly understood allocation mechanism, called the standardized multiple-choice aptitude test. We need more and better information, which is currently difficult to obtain, in order to understand these tests—because they function as the gatekeeping device to opportunities for making a decent living and contributing to society.

This present system of allocating opportunities is not benign. It is harmful. Every year thousands of competent students are unfairly denied opportunities to further their education, work and life skills. These falsely rejected individuals suffer as do the enormous numbers of potential clients of their services and skills.

The problem of allocating education and work opportunities with severely limited and secretive devices urgently needs to be opened to public scrutiny.

Test developers and users have been withholding facts about poor test construction, validation, and meaning within a shroud of innocence—and for far longer than most commercial institutions in our society have been allowed to do without being brought to an open accounting.

In business or industry, tests came under careful scrutiny in the 1960's after the Supreme Court ruling on *Griggs* v. *Duke Power Company.* Since then, industry has been made to demonstrate the validity, value, and reasonableness of its methods of selection—or work allocation—strategies.

We should not be tolerant of excessive secretness in testing individuals and making life decisions for them. The strength of truth-in-testing legislation is that it opens the testing process to scrutiny, understanding, public information and ultimately accountability by making known both the strengths and weaknesses of standardized testing. Tests would be subject to outside scrutiny and evaluation for the first time.

Truth-in-testing legislation should be only the first step in the demand for accountability. It should not become merely a placebo as Dr. Kenneth Clark recently warned in a letter to the New York Times.

In the next section I have written about some of the practical issues of the bill. I think Dr. Perrone and others have covered many of these points. The costs are not significant, they are exaggerated, I think to deflect attention from the real issue of openness and accountability.

The validity would probably not be affected with respect to predicting real life outcomes beyond academia, that is work outcomes. There is no evidence that the tests are valid predictors of anything of consequence beyond academia.

I want to focus some comments on the problem of not getting information and not knowledgeable to analyze the tests because of the testing industry's resistance to being held accountable.

The testing companies, concerned about the financial impact of selling their products to people who would be no longer in the dark about the value of these tests, will argue against this law. They want to discourage any sun from shining on test development, utilization and interpretation processes.

They present themselves as public servants who do not compete in business, in fact, call themselves testing agencies rather than companies as a way of disguising their revenue seeking motives.

Yet, ETS reported $1 million in excess revenue after expenditures in 1978—out of a total of $80 million in revenues.

I see nothing wrong with their seeking revenue, but if they cannot be open and above board about something so obvious as their desire to make money, how can we trust their statements about less obvious matters? Their statements breed mistrust and lead many of us not to believe them even when they are not lying to us.

We must look beneath their public relations policies and actions to understand their real positions. Recently, E. Belvin Williams, vice president of the Educational Testing Service, in a discussion about New York's truth-in-testing legislation, stated publicly on the television show "Today", and I quote, "I think that the law in New York State is a good one, although it created certain kinds of problems. Students should see their tests."

This was said publicly when in fact they launched and bankrolled an elaborate and extensive lobbying effort to defeat the New York bill. They are now working with equal vigor to rescind this bill which they publicly praise.

The point is test companies will tell you that many people have access to their tests. But in a case I was closely associated with, it took court action to be allowed to see a test that was disputed, and even then the courts agreed to a test company's stipulation that we only be allowed to view the tests for a limited amount of time.

The courts held up our ability to examine just fundamental, basic information about the tests for over a year.

The truth is that every aspect of this law is doable, practical, necessary, and overdue. But these clever and professional marketing experts will try to confuse and impress you with accounts of tests development costs, complications, and technicalities. These ploys are aimed at persuading you that disaster for their clients will occur if they are forced to be accountable.

The red herrings of higher costs, demise of validity, enforcement problems, or others put before you by the law's detractors, are smoke screens for the real issues of openness and accountability.

What about multiple-choice tests? How good are they? It is ironic that the very instrument of promise for fair (or equitable) and competence-based allocation of opportunities for education should now come under attack, but the truth is that we social scientists blew it. We reduced personal competence to trivial test-taking skills because that was the only practical technology we had available to us.

More valid and less discriminatory technology is available. It simply is not being used. The testing companies are locked into their old techniques. Rather than inform the public about new technology and persuade the public to make use of it, they continue to sell cheap tests that assess few variables of consequence, and they want us to believe that this is all, or the best, that is available and that it is sufficient. Neither is true.

Multiple-choice aptitude tests are unfair because they fail to assess the more critical aspects of competence that determine how effective students could be as workers and productive citizens. This means that we are not selecting people into education and work based on individuals' potential for competence on the job. Rather, we allocate education, work and life opportunities solely on the basis of some stripped-down version of intelligence—such a memory for facts or tests-taking ability, not even academic competency.

The ability to perform for a matter of minutes limited intellectual skills of recognition, recall, or esoteric analysis is an absurd way of allocating opportunities of lifelong importance. It is neither scientifically nor morally defensible to generalize from one single and simplistic test procedure to complex learning, working, and living experiences.

Decisions can be based on more than these limited measures of people's overall competence in life. But they won't be so long as we continue to believe in the myth of multiple-choice aptitude tests, and we will continue to believe in these myths as long as we are not provided more information about them.

Test companies advise admissions officers not to rely exclusively on test scores, but they also promote test score validity. Admissions officers often do not seriously question test validity. Or they use tests as convenient decisionmaking devices regardless of the validity, that is, they accept the utility of the tests without understanding them.

Accurate data and information about the meaning of these test scores is so hard to obtain that practically speaking, learning about the true value of tests is not worth the effort for most test users.

In short, our society's covenants with its citizens with respect to opportunities for education and work are determined unfairly and without merit on the basis of a few limited and irrelevant abilities to solve trivial problems. These abilities bear little relationship to the more enduring qualities or characteristics of productive and good citizens.

We must learn if tests are contributing to the loss of our ability to think, analyze, understand, and write because education stoops to cheap but convenient disguises for knowledge and competence. We must know if we are decreasing the intellectul nourishment of our youth by seducing them into memorizing facts or using phony analytical skills on trivial multiple-choice problems so they qualify for education and work opportunities.

Testing processes need to be opened to public scrutiny, understanding and evalution. Truth-in-testing legislation is urgently needed. Its requirements are doable. Its results will be enlightenment, more equity and attention to competence in allocating education and work opportunities, and yes, even better tests.

[The prepared statement of Dr. Pottinger follows:]

PREPARED STATEMENT OF PAUL S. POTTINGER, PH. D., DIRECTOR, NATIONAL CENTER FOR THE STUDY OF PROFESSIONS, WASHINGTON, D.C.

I. INTRODUCTION

My name is Paul S. Pottinger. I am the executive director of the National Center for the Study of Professions here in Washington. I am testifying before you today in my official capacity as the Center's director and spokesperson for Truth in Testing.

The Center was established nearly two years ago as a private not-for-profit organization. We were started with a grant form the M.A. Rooney Foundation in Illinois. Currently, our funds come primarily from the National Institute of Mental Health for testing and manpower research in the mental health profession.

Our mission is partly to raise the level of awareness of the public about issues of manpower regulation within the professions. We study certification and licensing practices among other regulatory mechanisms.

We view tests as a major regulatory device in the professions, and we are concerned about the effects tests have on the freedom of competent citizens to pursue work of their choosing.

Part of the Center's mission is also to help examine alternatives to currently used testing techniques. My research background is related to this. I have a doctorate in psychology from the University of North Carolina in Chapel Hill and a Bachelor of Science degree from Denison University in Ohio. I was Senior Research Associate in the Office of Research at the National Institute of Education (DHEW) from 1973-74. For three years prior to establishing the National Center for the Study of Professions I was the Director of Assessment

Systems at McBer and Company, a Boston social science research and development consulting firm.

I recenlty edited a book titled, Defining and Measuring Competence (Jossey-Bass, 1979). This book, including a chapter that I wrote, articulates some of the problems of current testing practices and the need for more and better information about tests.

So I come here today as both an advocate of proper test development and usage and as an expert. I am confident that better alternatives to our current testing practices will become possible if we can learn more about the tests at issue in these bills.

I am here today to support Truth-in-Testing legislation.[1] I think it is urgently needed because

- tests are a permanent part of our decision-making about who is given the opportunity to learn and to work,
- test validity, use and meaning are poorly understood by the public, whose lives are affected, and
- there are few if any incentives for making information abouth tests publicly available and better understood.

Accountability, control over our lives and basic human rights are at stake. Therefore, I want to address four important issues:

- The rationale for and urgency of passing Truth-in-Testing legislation.
- The practicality of this legislation in terms of costs, validity and enforcement.
- The testing industry's resistance to making information available or to being held accountable.
- The value of multiple-choice aptitude testing in terms of fairness and relevance to individual competence.

---

[1] I refer to truth-in-testing as a general concept. I believe, in fact that the Educational Testing Act of 1979 (H.R. 4949) is the stronger and more preferable of the two bills under consideration.

II. RATIONALE AND URGENCY OF PASSING TRUTH-IN-TESTING LEGISLATION

> **We allocate scarce education and work opportunities with aptitude tests. Few people understand these tests. Test developers and users do not provide information needed for us to understand. Yet we do not hold this industry accountable for their secretive probing of our minds. Current testing practices are harmful to this country's goals of equity and a competent work force.**

The problem we face today in classifying our youth and other citizens to allocate education and work opportunities is a crisis of proportions unimagined fifteen years ago, or even more recently. Today our country's work force requires more education than ever before; and jobs are scarce. Our ways of allocating scarce education and work opportunities have become more and more critical.

We are using a very poorly understood allocation mechanism, called the standardized multiple-choice aptitude test. We need more and better information, which is currently difficult to obtain, in order to understand these tests -- because they function as the gatekeeping device to opportunities for making a decent living and contributing to society.

This present system of allocating opportunities is not benign. It is harmful. Every year thousands of competent students are unfairly denied opportunities to further their education, work and life skills. These falsely rejected individuals suffer as do the enormous numbers of potential clients of their services and skills.[2]

---

[2] Education itself suffers as it is forced to mold its curriculum to meet the demands of a few test developers. Innovation suffers because any school that strays from the commercial tester's blueprints for learning will certainly pay a penalty for lower test performances in loss of support by alumni, the press and the public. All of these people, for lack of good information, believe in multiple-choice test scores as a matter of faith and ignorance. They equate test scores with individual and program competence.

The problem of allocating education and work opportunities with severly limited and secretive devices urgently needs to be opened to public scrutiny. Test developers and users have been withholding facts about poor test construction, validation, and meaning within a shroud of innocence -- and for far longer than most commercial institutions in our society have been allowed to do without being brought to an open accounting.

In business or industry, tests came under careful scrutiny in the 1960s after the Supreme Court ruling on Griggs v. Duke Power Co. Since then, industry has been made to demonstrate the validity, value and reasonableness of its methods of selection (or work allocation) strategies.

Nothing comparable exists in our system of allocating opportunities for the learning that is increasingly required for many jobs. So today, the problem of allocating scarce education and work opportuinties is a problem of enormous proportion. It centers around the use of a simplistic mechanism mechanism that needs more exposure, analysis and understanding by the public.

We should not be tolerant of excessive secretness in testing individuals and making life decisions for them. The strength of Truth-in-Testing legislation is that it opens the testing process to scrutiny, understanding,public information and ultimately accountability by making known both the strengths and weaknessess of standardized testing. Tests would be subject to outside scrutiny and evaluation for the first time.

Truth-in-Testing legislation should be only the first step in the demand for accountability. It should not become merely a placebo as Dr. Kenneth Clark recently warned in a letter to the New York Times (August 18, 1979).

III. THE PRACTICALITY OF TRUTH-IN-TESTING LEGISLATION

> **The cost of developing new tests would be insignificant. The validity of the tests, especially in predicting important life outcomes, would remain unaffected. The regulation of education or of test developers and users should not be confused with enforcement. This law is to make test publishers and users open and accountable.**

A potential problem is cost. More test questions will have to be written if each test is released. The cost of developing and using tests is not insignificant, but the cost incurred by writing more test questions is insignificant. In fact, a small percent of tesintg fees are spent on the writing of tests. Most of the cost of admissions testing is in the labor-intensive administrative aspects of testing -- not in development.

There are at least a dozen major technical steps in developing mulitple-choice tests, and one of the cheapest steps is item generation. The editing and testing of items is not expensive either. The latter step, for example, is often done by including new items in tests that are already paid for and in use. The added cost of writing new tests could be insignificant especially when spread across hundreds of thousands of tests.

The exaggerated concern about extra dollars required to enable students, researchers, teachers, college personnel or others to know on what basis life decisions are being made has become a tactic often used by test developers themselves to deflect attention from the real issues of openness and accountability.

The issue of validity is another concern. Will tests which might use items that have been made public still be valid? The answer is yes: Tests will be at least as valid as they have been.

The potential number of items for any of the tests that would be affected by this law is so great that no one could possibly memorize the answers to all the thousands of items in the item bank. (And if they did, they might know more than many who knew little but guessed well on the test.)

What is more important, these tests currently have no validity with respect to work and life skills required outside of academia. They are sometimes poor predictors of academic competence, but they are always poor predictors of competence in the real world.

Thus, many students are inappropriately screened out of opportuinites to enrich their existing work and life-relevant competencies because they lack some esoteric knowledge or skill valued only in academia.

For example, how would you feel about your son or daughter losing an educational opportunity because he or she could not answer the following types of questions to the satisfaction of the testers?[3]

IMPREGNABLE : AGGRESSION ::
(A) imperfect : revision
(B) invincible : defense
(C) inequitable : criticism
(D) indivisible : separation
(E) immutable : preservation

---

[3] These items are from the S.A.T., from Wheeler, Thomas C., "The American Way of Testing", The New York Times Magazine, September 2, 1979.

SWILL : SWINE ::
(A) roe : fish
(B) Coop : poultry
(C) mouton : sheep
(D) pesticide : vermin
(E) fodder : cattle

After reading a short passage about the ways of the jungle and the desert, you are asked questions in the following format:

Which of the following is (are) true of both the way of the jungle and the way of the desert?

I. They are characterized primarily by the struggle of creature against creature.

II. They are reactions to hostile environments.

III. They result in population control.

(A) II only
(B) III only
(C) I and II only
(D) II and III only
(E) I, II and III

Furthermore, what does the ability to answer these questions have to do with being competent at work and in life -- for which education presumably is preparing one? New law would make the lack of validity of these tests public information so that colleges, students and others could know when and when not to rely on tests.

Another concern is whether the feds will regulate educators or test developers and users. Regulation is neither the intent nor do I believe it will be the outcome of this bill.

Some mechanism such as civil procedures for preventive relief or administrative procedures must be required to enforce this law. But enforcing openness about testing activities is not the same as regulating the activities themselves. (If the specter of regulation is enhanced by the request for the Commissioner of Education to report on testing, the requirement could be dropped.)

IV. THE TESTING INDUSTRY'S RESISTANCE TO BEING HELD ACCOUNTABLE

> Testing companies have a large financial stake in selling multiple-choice tests. An informed public is threatening to their credibility. Some are slick marketers and defenders of their tests. They have many rationalizations for not being held accountable. The law is a sound one and overdue.

Testing companies, concerned about the financial impact of selling their products to people who would be no longer in the dark about the value of these tests, will argue against this law. They want to discourage any sun from shining on test development, utilization and interpretation processes.

The testing companies should fear loss of credibility and revenue if the truth gets out about tests. This is a legitimate fear based on their own knowledge that the tests we are talking about fail to be fair or competence-based.

Fear of accountability has already led to widespread hypocricy and deceit by some testing companies. For example, to persuade us that they are public servants rather than salespeople, some of them claim that they do not market their products or compete with one another for business. They even call themselves testing agencies rather than companies to disguise their revenue-seeking motives. Yet, ETS reported one million dollars in excess revenue after expenditures in 1978 -- out of a total of $80 million in revenues.

I see nothing wrong with their seeking revenue, but if they can not be open and above board about something so obvious as their desire to make money, how can we trust their statements about less obvious matters? Their statements breed mistrust and lead many of us not to believe them even when they are not lying to us.

We must look beneath their public relations policies and actions to understand their real positions. Recently, E. Belvin Williams, vice president of the Educational Testing Service in a discussion about New York's Truth-in-Testing legislation, stated publicly on the television show "Today" (and I quote) "I think that the law in New York State is a good one, although it created certian kinds of problems. Students should see their tests." This was said publicly when in fact they launched and bankrolled an elaborate and extensive lobbying effort to defeat the New York bill. They are now working with equal vigor to recind this bill which they publicly praise.

Test companies will tell you that many people have acess to their tests. But in a case I was closely associated with, it took court action to be allowed to see a test that was disputed, and even then the courts agreed to a test company's stipulation that we only be allowed to view the tests for a limited amount of time. No notes could be taken.

When we asked for validity information we received numbers that were patently false. When we pointed out to the testing company that their numbers could not have been right because of arithmetic errors they claimed the problem was just a matter of "typos" and immediately retracted the data offered as evidence for their case.

The truth is that every aspect of this law is doable, practical, necessary and overdue. But these clever and professional marketing experts will try to confuse and impress you with accounts of test development costs, complications and technicalities. These ploys are aimed at presuading you that disaster for their clients will occur if they are forced to be accountable. The red herrings of higher costs, demise of validity, enforcement problems, or others put before you by the law's detractors, are smoke screens for the real issues of openness and accountability.

V. THE VALUE OF MULTIPLE-CHOICE TESTING

> Admissions aptitude tests are unfair because they reduce personal competence to memory for facts, esoteric analysis or test-taking skills. Other competencies critical to work and life are not assessed. These tests are unfair because they contain measurement error and because this very fact is frequently of no consequence to test users. These tests are social sorting mechanisms, unworthy of current social science and social values.

What about muliple-choice tests? How good are they? It is ironic that the very instrument of promise for fair (or equitable) and competence-based allocation of opportunities for education should now come under attack, but the truth is that we social scientists blew it. We reduced personal competence to trivial test-taking skills because that was the only practical technology we had available to us.

As a result few if any of standardized testing's promises have been achieved. Rather we have created a multimillion dollar industry that must perpetuate the technology of multiple-choice tests.

More valid and less discriminatory technology is available. It simply is not being used. The testing companies are locked into their old techniques. Rather than inform the public about new technology and persuade the public to make use of it, they continue to sell cheap tests that assess few variables of consequence, and they want us to believe that this is all (or the best) that is available and that it is sufficient. Neither is true.

Multiple-choice tests are not fair because they contain large measurement error (as much as 60 points on some tests). Thus numerous people are excluded from opportunities by measurement error, often in their one shot at opportunity.

Multiple-choice aptitude tests are unfair because they fail to assess the more critical aspects of competence that determine how effective students could be as workers and productive citizens. This means that we are not selecting people into education and work based on individuals' potential for competence on the job. Rather we allocate education, work and life opportunities solely on the basis of some stripped-down version of intellignece -- such as memory for facts or test-taking ability.

The ability to perform for a matter of minutes limited intellectual skills of recognition, recall or esoteric analysis is an absurd way of allocating opportunities of life-long importance. It is neither scientificaly nor morally defensible to generalize from one single and simplistic test procedure to complex learning, working and living experiences.

Decisions can be based on more than these limited measures of people's overall competence in life. But they won't be so long as we continue to believe in the myth of multiple-choice aptitude tests, and we will continue to believe in them as long as we are not provided more information about them.

The tests are unfair because they are sometimes invalid for predicting academic competence. One test expert from a major aptitude testing company recently told me that he wanted to tell some schools about the validity (for them) of a college aptitude test. The validity was so low for some schools that it was essentially meaningless. But he feared for his job if he told the truth; his company might lose the school's business. This anecdote reveals the motivation of the testing companies to sell their products regardless of their real value; it also reveals the lack of information available to the schools to make their own determination of the value of the tests for themselves.

Test companies advise admissions officers not to rely exclusively on test scores, but they also promote test score validity. Admissions officers often do not seriously question test validity. Or they use tests as convenient decision-making devices regardless of the validity, i.e., they accept the utility of the tests without understanding them. Accurate data and information about the meaning of these test scores is so hard to obtain that practically speaking learning about the true value of tests is not worth the effort for most test users.

We are enchanted with these aptitude measures and the mystery surrounding them as we were some years ago by IQ tests. Indeed we can trace the origins of our widely embraced use of multiple-choice aptitude tests to conditions surrounding the advent of IQ tests.

IQ tests became popular because they served a need to classify people not because they were scientifically sound. The inappropriateness -- even the immorality -- of using IQ tests to track and sort children is now widely recognized and accepted.

So why are we so enamored with aptitude tests? Because they too are convenient sorting tools. But the test developers must fight against our realizing the similarities between aptitude tests and other measures which many educators and businesspeople have already discarded as useless or harmful.

Only recently has there been widespread attention to the inherent discriminatory effects of these tests, which reflect and perpetuate the social philosophies of the dominant social groups that design them.

Current testing practices must be evaluated in light of both practical classification needs and our philosophical beliefs about social contracts and human dignity. Our social values have changed. Status in our society is changing from being based on intelligence tests or academic achievement to being based on demonstrated competence. Our tests simply do not measure up to this change in our values and perspectives. We have put the old wine of IQ tests into new bottles and found it erringly more tasteful.

VI. SUMMARY

In short, our society's covenants with its citizens with respect to opportunities for education and work are determined unfairly and without merit on the basis of a few limited and irrelevant abilities to solve trivial problems. These abilities bear little relationship to the more enduring qualities or characteristics of productive and good citizens.

We must learn if tests are contributing to the loss of our ability to think, analyze, understand, and write because education stoops to cheap but convenient disguises for knowledge and competence. We must know if we are decreasing the intellectual nourishment of our youth by seducing them into memorizing facts or using phony analytical skills on trivial multiple-choice problems so they can qualify for education and work opportunities.

Tesing processes need to be opened to public scrutiny, understanding and evaluation. Truth-in-Testing legislation is urgently needed. Its requirements are doable. Its results will be enlightenment, more equity and attention to competence in allocating education and work opportunities, and (yes) even better tests.

We should not allow ETS and other members of education's mafia to control our access to needed information. We need sunshine and openness in testing. I sincerely hope a strong testing bill will be passed by Congress.

Chairman PERKINS. Thank you for excellent testimony.

We are going to revert to the first witness. It would be my hope that the members would take about 7 or 8 minutes and perhaps that will give you time to ask each witness a question. When we come back to the second go-around you can run 15 or 20 minutes. None of the members will be cut short.

Mrs. Sparling, how would you answer the criticism of the Weiss bill that it would lead to an undue burden on the testing industry and increase costs to students since a test could only be used once before it was made public?

Mrs. SPARLING. Mr. Chairman, we have discussed this morning the number of questions in the test bank so we would have to have more information as to how the bill is related to public disclosure. We would need information to determine that question. We are primarily interested in how these tests are going to affect our young people.

Chairman PERKINS. Now, Dr. Rever, how do you answer the criticism that the testing industry, which is so important to the future job prospects of students, is basically an unregulated industry not accountable to the public?

Dr. REVER. Mr. Chairman, the test industry as you have witnessed here before you have two segments in general. They have the for-profit segment regulated by existing law both State and local. The nonprofit sector which I represent, which is represented by the Educational Testing Service, is to some degree regulated by the Code of the Internal Revenue Service.

But with regard to specific activities as reflected in Mr. Weiss' bill, we find that the kinds of issues that are being addressed in this bill are addressed by the colleges and the universities that provide elected representatives for our governance. In fact, we are accountable to elective representatives of the higher education and the secondary education system in this country. They are the ones that make sure that we can state to you specifically that we see no need for this kind of legislation.

Chairman PERKINS. Dr. Perrone, how would you answer the criticism of the Weiss bill raised by the Medical and Dental Associations that there is not an adequate pool of questions for them to formulate a new test every year or several times a year since they would have to make public any test after it is given?

Dr. **Perrone**. I think I tried to argue already, Mr. Chairman, that the item pool is very, very large. There is absolutely no evidence that I am aware of, that the MCAT or the examinations used for that process deal with a limited base of information. In fact, the MCAT has a very large section that talks about problem-solving skills and the like. It is almost infinite in the terms of the numbers of questions that could be devised to respond to that.

I think it might also be interesting for the committee to note that in regard to the MCAT what little bit of research there has been and correlations between the MCAT and success in medical school and success later, the Weingard-Williams research that covers a period of about 10 years indicates that the MCAT does not correlate effectively at all with college grades and medical school, success in medical school or in clinical study or with success as a practitioner.

Chairman **Perkins**. Let me ask Dr. Snyder a question.

Doctor, regarding occupational testing, do you believe adequate safeguards now exist to assure that test-takers are fairly treated? Congressman Gibbons from Florida does not believe such safeguards now exist.

What is your answer?

Dr. **Snyder**. Yes, I do. I believe that prior legislation, especially around the Equal Opportunity Act, adequately set a standard and a stage for demanding that validity evidence be available for any anticipated tests use. I believe that is adequate and built into the disclosure provisions and the necessity of preserving the privacy of individuals being tested is also adequately covered by the other Privacy Act.

CONTROL OF TESTING INDUSTRY

Mr. **Millenson**. May I comment on a previous question?

Chairman **Perkins**. Yes.

Mr. **Millenson**. I would like to comment on the question posed with respect to control of the testing industry. Except for college entry testing, the greatest part of the industry's tests are sold to local and State educational agencies and bodies. Therefore, we are dealing with local and State governments. If there is to be any control over the testing industry, then it would have to be at the Federal level with respect to purchases which are made by local and State governments which I think raises some very serious questions with respect to the Federal controls of education, let alone with respect to the adequacy of the U.S. Office of Education controlling such an item.

Chairman **Perkins**. Thank you very much.

Now, Dr. Pottinger, if I understood you correctly, you stated you were an expert. Since your center studies all the professions, let me ask you about a criticism of both bills.

Do you believe that the problems asserted about testing are grave enough to have the Federal Government come in and preempt the States in this area?

Dr. POTTINGER. Yes, Mr. Chairman, I do. I think in fact in response to a remark just made by Dr. Snyder, that EEOC is adequate protection for the occupations, we find in fact that the testers argue that EEOC does not apply to certification and licensing tests. In fact, it has not been tested in the courts. The testing companies tell us they will proceed without abiding by EEOC guidelines until there is a test case in the courts and I believe that that attitude by the testing companies that they will do the minimum necessary until forced by the courts to do more, reflects in my opinion the need to adequately oblige them with the kind of enforcement procedure required to make tests accountable.

Chairman PERKINS. Mr. Goodling.

Mr. GOODLING. Thank you, Mr. Chairman.

I have several objections but perhaps most of my questions will wait for the second round.

First of all, my major concern about this whole legislation puts me in an awkward position. As a counselor, principal, superintendent, and supervisory teacher for 23 years, I was always very concerned about the relationship of how much standardized testing we did and what the results were used for.

I want to make very sure we don't go off half cocked. I think when there is an emergency and we pass some legislation, and between the legislation and the regulations we don't help the people that we were trying to help, perhaps we can be forgiven. When there is no emergency, as is true in this case, then we certainly could not be forgiven if in fact we just end up doing opposite of what we thought we were going to be doing in the first place.

For instance, when we had the sponsors of the legislation testify before the subcommittee, the response that I got I think from the one gentleman sponsoring one of the bills was its purpose was pretty much centered around whether someone did or did not get into a law school.

In the second piece of legislation it appeared that the major concern dealt with—and I am not sure this is correct—helping young people who are disadvantaged. Because they are disadvantaged, they don't have the money for tutoring, et cetera.

It certainly seems to me, if we are ever going to hurt those disadvantaged individuals, we will hurt them with this legislation. We'll create the kind of opportunity for private groups to get in the business of tutoring more than they are today unless there is a companion bill, which will provide Federal funds to make sure that all have an equal opportunity to be tutored in order to take the tests.

So I have some real concerns about that. I don't know if you saw the editorial this morning in the Post, but I thought Mr. William Raspberry, who is certainly a noted columnist for the Post, makes several good objections.

I will just mention three quotes:

> One bill, introduced by Rep. Sam Gibbons (D.-Fla.) is legislative cotton candy. It looks good and tastes good, but it is mostly air. Its provisions fall into two main categories: the unnecessary and the impossible.

A second quote:

If the Gibbons bill is cotton candy, the other proposal—introduced by Reps. Ted Weiss (D.-N.Y.), Shirley Chisholm (D.-N.Y.) and George Miller (D.-Calif.)—is legislative horehound: strong, bitter stuff that doesn't cure anything.

Getting to a point in part of the testimony, and I quote:

> But since publication of any particular tests would lead to its immediate retirement, the corrected exams would be helpful only in the same generalized way as the test study guides already available.
>
> Indeed, as Diane Ravitch of Teachers College, Columbia University, has pointed out, the major benefit of test-answer publication would accrue not to those who fail the tests but to those who sell test-taking assistance. "It will create a bonanza for private tutoring services, which would have actual tests as teaching tools, not just their own inventions," she said. "This will increase the advantage of those who can afford to pay for coaching."

So I want to make sure as we go through this that we don't jump off the deep end rapidly and end up hurting the very people we are trying to help. It seems to me in the last couple years every time we take a potshot at education we refuse to consider some of the real problems. We refuse to consider the deterioration of family life, the instability at home, the competition for students' time. I just watched two teenagers grow up. They would have needed 48 hours to do all of the things that are expected of them to do in school and home and in the community and so forth. We refuse to accept the fact that in the last 15 years we have educated all students which is totally different than what it had been prior to that.

First of all, if I have a few minutes left, I think I noticed in your testimony, Mrs. Sparling, you had a real concern about teachers teaching for the tests. Do you not feel that publishing these tests and the results and answers isn't going to make that more likely than it presently is and Lord knows it is being done now? Teachers feel insecure. They want to make sure when you stick a standardized test in front of them their youngsters do well.

Mrs. **Sparling**. With this particular bill there will be more information of the purposes and uses of the test so that the results of the tests were known to the students and parents then it could be used in a diagnostic manner so that the education of that particular student would be enhanced by this kind of knowledge.

Then it would also give the teacher the opportunity to take an overall look at the needs of that student, so that a better curriculum could be devised.

Mr. **Goodling**. Do you not find a trend in this country that these standardized tests are being used more diagnostically than anyting else at the present time other than in some select schools—perhaps the top 10 or 12?

Mrs. **Sparling**. This particular bill will be one that will give us this kind of information and then they certainly can be used to be more diagnostic.

Mr. **Goodling**. Do they not already have that opportunity?

Mrs. **Sparling**. But the parents and the students do not have the opportunity to see the results other than the raw scores.

Mr. **Goodling**. I see. This legislation is going to pull the family unit back together and make parents become more involved in the education of their children.

Mrs. **Sparling**. Parents are very interested in being involved in the decisions that affect the lives of their children and certainly

501

where education is concerned they are very concerned with the issue of testing.

Chairman PERKINS. Mr. Weiss.

Mr. WEISS. Thank you very much.

Just so that there is some balance of newspaper reporting—and this is not by way of editorializing—some questions have been raised as to the coaching schools, and the New York Times, in testing an education center, had a feature story. Let me just read one aspect of it. The Nation's largest test preparation school has branches in 88 cities. Then it goes on to quote about other schools. They were not talking abcut just a random situation of the testing phenomenon being used by profitmaking ventures to the value of those who could afford the schools and to the disadvantage of those who can't.

Dr. Perrone, first let me reemphasize the disclaimer you make that your statement is not to be viewed as an official position of the University of North Dakota, the North Dakota Study Group or the National Consortium on Testing. I appreciate and recognize that.

I do want to ask: You are the president of the National Consortium of Testing; is that correct?

Dr. PERRONE. That is correct.

Mr. WEISS. We have some indication as to the validity and standing of that group. I will ask permission to have entered into the record a list of members of the National Consortium of Testing as of June 1 of 1979.

Chairman PERKINS. Without objection.

[The information referred to above follows:]

NATIONAL CONSORTIUM ON TESTING—LIST OF MEMBERS

June 1, 1979

1. American Association of Physics Teachers, Graduate Physics Building, SUNY at Stony Brook, NY 11794, (516) 246-6840. Designated Representative: Edwin F. Taylor, 86 Oxford Road, Newton Center, MA 02159.
2. American Association of School Administrators, 1801 North Moore Street, Arlington, VA 22209, (703) 528-0700. Designated Representative: William G. Spady.
3. American College Testing Program, P.O. Box 168, Iowa City, Iowa 52240, (319) 356-3845. Designated Representative: Philip R. Rever, ACT Washington Office, One Dupont Circle, N.W., Suite 340, Washington, D.C. 20036, (202) 223-2318.
4. American Educational Research Association, 1230 17th Street, N.W., Washington, D.C. 20036, (202) 223-9485. Designated Representative: Peter W. Airasian, School of Education, 600 McGuinn, Boston College, Chestnut Hill, MA 02167, (617) 969-0100.
5. American Federation of Teachers,[1] 11 Dupont Circle, N.W., Washington, D.C. 20036, (202) 797-4400. Designated Representative: Eugenia Kemble.
6. American Psychological Association, Office of Scientific Affairs, 1200 17th Street, N.W., Washington, D.C. 20036, (202) 833-7600. Designated Representative: Richard E. Snow, School of Education, Stanford University, Stanford, CA 94305.
7. American School Counselor Association, 1412 North 13th Street, Boise, Idaho 83702, (208) 345-6722. Designated Representative: Helen R. Washburn.
8. Association of American Medical Colleges, Division of Educational Measurement and Research, One Dupont Circle, N.W., Suite 200, Washington, D.C. 20036, (202) 466-5100. Designated Representative: Robert L. Beran.
9. Association for Childhood Education International,[1] 3615 Wisconsin Avenue, N.W., Washington, D.C. 20016, (202) 363-6963. Designated Representative: Elvie Lou Luetge.

---

[1] Designates members of the Board of Directors.

10. Association for Measurement and Evaluation in Guidance, Department of Measurement and Statistics, College of Education, University of Maryland, College Park, MD 20742. Designated Representative: William D. Schafer.
11. Association for Non-White Concerns in Personnel and Guidance, c/o William E. Gardner, President, Chairman, Dept. of Education, Lincoln University, Lincoln University, PA 19352, (215) 932-8300. Designated Representative: Mildred R. Buck, 10315 Corbeil, Apt. A, St. Louis, MO 63141.
12. Association for Supervision and Curriculum Development, 225 No. Washington St. Alexandria, Va. 22314, (703) 549-9110. Designated Representative: Ruth Long.
13. Center for Applied Linguistics, 1611 North Kent Street, Arlington, VA 22209, (703) 528-4312. Designated Representative: Roger Shuy.
14. Chicano Education Project, 1444 Stuart Street, Denver, CO 80204, (303) 825-7364. Designated Representative: Bill Rosser.
15. Child Welfare League of America, Inc., 1346 Connecticut Avenue, N.W., No. 310, Washington, D.C. 20036, (202) 833-2850. Designated Representative: William L. Pierce.
16. The College Board, 888 Seventh Avenue, New York, NY 10019, (212) 582-6210. Designated Representative: Stephen H. Ivens.
17. Community Resources Institute, 670 West End Avenue, No. 12F, New York, NY 10025, (212) 595-0295. Designated Representative: Ann Cook.
18. Council for American Private Education, 1625 Eye Street, N.W., Washington, D.C. 20006, (202) 659-3236. Designated Representative: Robert Lamborn.
19. Council for Educational Diagnostic Services, c/o Louise Appell, National Committee, Arts for the Handicapped, 1701 K Street, N.W., Suite 801, Washington, D.C. 20006, (202) 223-8007.
20. Council for Exceptional Children,[1] 1920 Association Drive, Reston, VA 22091, (703) 620-3660. Designated Representative: Jean N. Nazzaro.
21. Division for Study and Research in Education, Massachusetts Institute of Technology (MIT), Cambridge, MA 02139, (617) 253-2050. Designated Representative: Judah L. Schwartz.
22. Education Commission of the States,[1] 1860 Lincoln Street, Suite 300, Denver, CO 80295, (303) 861-4917. Designated Representative: Homer Elseroad.
23. Education Development Center,[1] 55 Chapel Street, Newton, MA 02160, (617) 696-7100. Designated Representative: Jerrold R. Zacharias.
24. Educational Records Bureau, 2 Sun Life Executive Park, 100 Worcester St., Wellesley Hills, MA 02181, (617) 235-8920. Designated Representative: R. Bruce McGill.
25. Educational Testing Service, Princeton, NJ 08540, (609) 921-9000. Designated Representative: Alice J. Irby, ETS, 11 Dupont Circle, N.W., Washington, D.C. 20036, (202) 483-5110.
26. International Reading Association, 8000 Barksdale Road, Newark, DE 19711, (302) 731-1600. Designated Representative: Roger Farr, Indiana University, Education Building, Room 227, Bloomington, IN 47401, (812) 337-1440.
27. National Association for the Advancement of Colored People,[1] 1790 Broadway, New York, NY 10019, (212) 245-2100. Designated Representative: Althea T. L. Simmons.
28. National Association for the Education of Young Children, 1834 Connecticut Avenue, N.W., Washington, D.C. 20009, (202) 232-8777. Designated Representative: Marilyn M. Smith.
29. National Association of Elementary School Principals,[1] 1801 North Moore Street, Arlington, VA 22209, (703) 328-6000. Designated Representative: Paul L. Houts.
30. National Association of Independent Schools, 4 Liberty Square, Boston, MA 02109, (617) 542-1988. Designated Representative: Wellington V. Grimes.
31. National Association of School Psychologists, c/o Ann Engin, President, 3781 Criswell Drive, Columbus, OH 43220, (614) 422-8787 (office telephone).
32. National Center for the Study of Professions, 1527 New Hampshire Avenue N.W., Washington, D.C. 20036, (202) 232-2204. Designated Representative: Paul S. Pottinger.
33. National Congress of Parents and Teachers, 700 N. Rush Street, Chicago, IL 60611, (312) 787-0977. Designated Representative: Ann Kahn, 9202 Ponce Place, Fairfax, VA 22030, (703) 591-4589.
34. National Council of Teachers of English, 1111 Kenyon Road, Urbana, IL 61801, (217) 328-3870. Designated Representative: John C. Maxwell.
35. National Council on Measurement in Education, 103 Draycott Street, Fayetteville, NY 13066 (home), (315) 423-3024 (office). Designated Representative: Eric F. Gardner.

36. National Education Association, 1201 Sixteenth Street, N.W., Washington, D.C. 20036, (202) 833-4000. Designated Representative: Frances S. Quinto.
37. National Employment Counselors Association, c/o Department of Commerce, Employment Security Division, 303 Caldwell Building, Tallahassee, Florida 32304, (904) 488-6005. Designated Representative: Richard O'Mara.
38. National School Boards Association,[1] 1055 Thomas Jefferson Street, N.W., Washington, D.C. 20007, (202) 337-7666. Designated Representative: James A. Mecklenburger.
39. National Science Teachers Association, 1742 Connecticut Avenue, N.W., Washington, D.C. 20009, (202) 265-4150. Designated Representative: Phyllis Marcuccio.
40. National Vocational Guidance Association, c/o Daniel Sinick, President, Professor of Education, Rehabilitation Council Education Program, George Washington University, Washington, D.C. 20052, (202) 676-6376. Designated Representative: Juliet V. Miller, 8751 Burneth Drive, Milan, Michigan 48160.
41. New York Public Interest Research Group, 5 Beekman Street, New York, NY 10038, (212) 349-6460. Designated Representative: Stephen Solomon.
42. North Dakota Study Group on Evaluation,[1] Center for Teaching and Learning, University of North Dakota, Box 8158, Grand Forks, ND 58201, (701) 777-2674. Designated Representative: Vito Perrone.
43. Public Education Association, 20 West 40th Street, New York, NY 10018, (212) 254-6100. Designated Representative: David Seeley.
44. Social Policy Institute, 33 West 42nd Street, Room 1212, New York, NY 10036, (212) 840-1258. Designated Representative: Frank Riessman.
45. Southeastern Public Education Program, American Friends Service Committee, 401 Columbia Building, Columbia, SC 29201, (803) 252-0975. Designated Representative: M. Hayes Mizell.
46. Speech Communication Association, State University College at Brockport, SUNY, Brockport, New York 14420. Designated Representative: Fred E. Jandt.
47. Teachers of English to Speakers of Other Languages, 455 Nevils Building, Georgetown University, Washington, D.C. 20057, (202) 337-7264. Designated Representative: James E. Alatis.
48. Workshop Center for Open Education,[1] Shephard Hall, Room 6, City College Advisory Service, Convent Avenue and 140th Street, New York, NY 10031, (212) 368-1619. Designated Representative: Lillian Weber.

CONSULTANT TO THE BOARD

Professor Banesh Hoffman, 43-17 169th Street, Flushing, NY 11358, (212) 358-6231.

Mr. WEISS. Again recognizing that not all of the organizations that are members of the National Consortium, would agree with either your statement or with the proposed legislation which I sponsor, let me just indicate again for credibility purposes only as to your credentials and that of the National Consortium itself, that it includes, among others, the American Federation of Teachers, the American Psychological Association, the American College Testing Program, the American School Counsellors Associations, the Association of American Medical Colleges, the College Board, the Educational Testing Service and so on.

So that this National Consortium of Testing is a very highly regarded organization in its field, I assume. Is that correct?

Dr. PERRONE. It is made up of a large number of organizations interested in a lot of issues that this legislation addresses.

Mr. WEISS. Would you again very briefly summarize what is the work of the National Consortium?

Dr. PERRONE. The National Consortium on Testing has as its primary purpose to increase public understanding of testing and tests use in this country, to bring about greater understanding of strengths and weaknesses of tests as they exist and is committed to assist in improving the quality of assessment practices that go on in our society in schools and outside of schools.

Mr. Weiss. Again to reemphasize, your position, as is the position of the sponsors of H.R. 4949, is not to eliminate tests but to provide more—supposing you state what your position is.

Dr. Perrone. The National Consortium obviously would not be supportive of eliminating the use of tests in our society. Clearly the aim of the group is to improve the quality of assessment practices that exist with and "toward greater fairness and equity across all segments of the society."

In my own case I am personally concerned that our present testing practices do not give support to a broad enough range of areas. I am concerned that alternatives to the current technology have not yet been explored fully because the current technology has not been fully exposed to the sunshine, as Paul Pottinger describes it.

I am certainly not opposed to standardized testing if applied fairly.

Mr. Weiss. We have had in the course of the hearings that the subcomittee has held some indication as to how available information about tests is currently and that this legislation really is not needed because anybody who wants the information, particularly experts in the field, would have no difficulty in getting all the available information.

What has been your experience as to the availability of information about the tests from the various testing services?

Dr. Perrone. There is clearly a limitation, not only a limitation for qualified scholars but I think a limitation more directly for those who must take those tests and those whose lives are governed by them.

I think what you might want to do is to enter into the record also Oscar Buros' valedictory address at the University of Iowa in 1972. Oscar Buros, who has been the noted chronicler around, evaluator of tests in this country over a 40-year period, who has been the prime producer of the Mental Measurement Handbook, in that valedictory statement made clear that not only were the tests very bad in the 1920's; he suggested to his great sorrow they were no better in the 1970's.

He went on further to say even though there is that claim that information is available to those who must use it for evaluation purposes, he stated quite explicitly that the materials were just never available to do the kind of assessment that was necessary if in effect the most critical questions about equity were to be responded to.

I think that statement would also be a very good entry into the record of this committee.

Mr. Weiss. Do you have that statement available?

Dr. Perrone. Yes, I can make that available to you.

Mr. Weiss. If you do, I think that would be well to have it in the record.

In Dr. Rever's testimony he stated, and I quote on page 8:

> The probable lack of recognition that different types of tests are constructed to serve diverse purposes will—this to me is the significant point—likely result in invalid interpretations and critiques of the items by individuals unschooled in testing methodology and uninformed about the intended uses of the tests.

Would you, Dr. Perrone, feel that may in fact in some variation summarize the concern that the testing services have not necessarily that there would be invalid criticism but that coming from recognized experts in the field there may be valid criticisms which may tend to suggest changes that should be built into the tests in order to make them fair.

Dr. **Perrone**. My sense is that any time you expose the test items, one, you are going to improve the opportunities for those tests to be made better and at the same time you are going to improve the opportunities for a lot of people to scrutinize them in ways that test publishers might not like and in fact experts in testing might find absurd.

But in the public exposure, the benefits to be gained are far greater than the deficits. Mr. Rever is essentially arguing that the deficits are going to be highlighted. My position tends to be that the opportunities for productive change are going to be highlighted.

Mr. **Weiss**. There has also been some suggestions that if you make tests available—and I would like both you and Dr. Pottinger to respond to this—to the students who took them and, of course, then to the public at large, there will be such a requirement for additional tests that it will be impossible, really, to supply the number of tests that are needed or that the costs will be so burdensome that the students will be the ones to suffer.

Dr. **Perrone**. The information that I have seen would suggest there is absolutely no validity to any of those claims. Now I must say there may be data somewhere hidden that is not available for the public record that I don't know about or that others in this room don't now about. But everything that I have seen would suggest that that claim has virtually no validity.

Mr. **Weiss**. Dr. Pottinger.

Dr. **Pottinger**. I would only add that there are numerous steps in the development of tests and the generation of new test questions, I think, as Dr. Perrone has already made known, is a very small aspect of this process. The fact is that many times items are developed by people in academia or by volunteers. The actual physical generation of test items is insignificant. The items then become edited and statistical procedures take place usually as the tests are being used by people who pay for them anyway, to examine their relationship to the tests characteristics and so forth.

All I can say is that both logically, procedurally and in fact the evidence that we have gathered from ETC is this is not a significant cost factor.

Chairman **Perkins**. Mr. Erdahl.

Mr. **Erdahl**. Thank you, Mr. Chairman.

A couple of observations and a question to any member of the panel who might wish to respond.

Ideally, for one in school, the perfect school would be one in which there were no tests! Life is full of tests, and schools are full of tests, and we are tinkering about, reducing the use and effectiveness of testing, so I have some reservations about the bill.

Are we looking for a serial number of standardized tests? It seems to me obviously as the test results are made available to the student who took the test, they are also available to other people who would be taking the tests. It seems to me we have a consensus here that the present testing system needs to be improved and modified to make it more fair, to make it coincide better with the student's actual aptitude and abilities.

But I am not so sure that this will really get it done.

By making this information available, will it in fact make the testing process better?

Are we assured under a new system that we wouldn't also see multiple choice questions being used which have been used by teachers before we had the standardized tests, so these are some questions we have not had answered yet.

Certainly, as Members of Congress, we use test scores for the appointments that we might recommend to the Military Academy. Are we hoping for a situation that no longer exists of having a more personalized way of evaluating people?

Ideally, I hope that we could do that, but it seems to me, as Mr. Goodling mentioned, we in fact could end up hurting the very people we seem concerned about trying to help, those that might be disadvantaged.

Do we make the possibility greater for tutoring situations, that wouldn't be available to the poor students? These are questions that certainly would have to be answered before I would support this bill.

Later on I will have some more specific questions.

What assurances do we have that this disclosure will really make the testing process better? Are we assured that under a new testing process we wouldn't have a repetitive new standardized test, that we would have to be turning over much quicker?

What assurance would we have that people would be accepted to a university, to a law or medical school, or some job opportunity by the fact that a person who took the test before knows the answers and therefore makes it available to anybody who takes the test?

TESTING ITEM POOLS

Dr. **SNYDER.** Basically, one very common thing that underlies your concerns whether or not the item pools will be large enough to generate the needed security in an individual administration, while still being able to release each and all of the items after they have been used at once.

I suspect that given the range of alternatives which are available with current technology, which is essentially to write more items or do a relatively speculative algorithm that will generate classes of items which will tend to make the contents of the test less unsuitable rather than more unsuitable.

Any time you are in a situation where you must generate an enormous range of contents and limit what you are measuring, that the specific questions used will get more trivial rather than less trivial.

Mr. **GOODLING.** I would like to comment on something that you said, perhaps.

You talked about tests given for admittance into the military academies and judgments that Members of Congress make.

As we know, a goodly number of Members of Congress utilize such tests themselves; and these are standardized tests, and the question should be posed how many Members, even of this committee, utilize these standardized tests, and how many members of this committee would be willing to have these very tests utilized in the method suggested by the pending bills? That is, give the answers.

How many of them are doing that now? They are worried about all of those other tests.

I would ask also the question, and I don't have the information as to the tests given in the House of Representatives for potential employees.

They give tests. I know they give typing tests and stenographic tests, but whether they give any other tests, because as we try to point out in our testimony, there is a double standard where we are asking the private sector, both profit and nonprofit, to do certain things which not only the Government does not do, but as you see to point out, even the members of this committee, the House of Representatives of the United States, doesn't have one.

Mr. **Erdahl**. If I could make another point I think most of us saw the article in the Washington Post. Another article, in other papers throughout the country, dealing with the SAT scores, was brought to my attention this morning. Someone mentioned, evidentally we see a decline in reading and writing skills, and yet maybe the problem is not with the testing but with the teaching procedure.

The tests, whether we like them or not, are fairly accurate, in that they show a decreased stability in these areas, and we shouldn't really blame the test for that. It's the procedures.

It's kind of like shooting the messenger of bad news. I am not sure the tests are the fault of the decline in some of these skills.

Chairman **Perkins**. Mr. Miller?

Mr. **Miller**. Thank you, Mr. Chairman.

Dr. Rever, on page 6 of your testimony, on item 2 you talk about the release of the securities items, and down at the bottom you say the tests will be inappropriately leading the curricula. A related effect would be to encourage the proliferation of coaching schools, many of which will fail to engage in sound educational practices.

Since it's not established that the efforts of coaching schools produce results of equal benefit to all and may vary for many students, and this would also be an anathema to the public interest.

I don't quite follow you there. You are suggesting if you release all these items, that then the public school system will engage in teaching just to pass the test?

Dr. **Rever**. The first point to be made in responding to your question is this: The college admissions test is for a purpose unrelated to the evaluation of the student's total experience in the high school, so consequently while Mr. Erdahl——

Mr. **Miller**. What do you mean?

Dr. **Rever**. Mr. Erdahl references a scholastic aptitude test and decline in test scores over time. We experience the same thing on the assessments, but we never stated that that was an indictment against the secondary school system in this country, because the content of the examination does not represent the entire wealth of the curriculum in the high school.

Mr. **Miller**. What does it represent then?

508

Dr. **Rever**. It represents what we believed what our users or when I say users, those are the secondary schools and the postsecondary institutions in this country, are the necessary abilities to survive in institutions of higher education in an academic sense.

We are in fact indeed——

Mr. **Miller**. Wait a minute; what is high school about?

Dr. **Rever**. High school, the curricula of the high school are used as the basis for the kinds of questions that we ask, but we do not intend to sample or draw questions from all the range of the high school curricula in this country.

Mr. **Miller**. The test will then provide, separate from your total high school experience, whether or not you have the skills to survive in an institution of higher education.

Dr. **Rever**. Our intent in developing the ACT assessment is to provide an instrument to predict students' performances in their first year in college, yes, indeed.

Mr. **Miller**. I thought that is what graduation from grades in high school was all about.

Dr. **Rever**. No; I don't believe so, as a matter of fact, one of the difficulties.

Mr. **Miller**. Why do we consider them, the grades, why do colleges consider the grades and the scores?

Dr. **Rever**. For two reasons; one is the grades among high schools, and according to the size of the high school and its curricula and the like, varies enormously in terms of their reliability in predicting first year in college performance in grades.

The test really is a third party external approach to providing some kind of opportunities for students to demonstrate the kinds of skills and abilities they have to an institution of higher education, independent of whatever the high school itself has tried to impart for the student.

Mr. **Miller**. I hate to say this, but I am still a little lost.

I thought if you went to high school and were a B student and you did B work in mathematics and algebra and trigonometry, your chances of doing fairly well in mathematics in college would be good.

If you did reasonably well in history in high school, you would do reasonably well in history in college, and the same as to literature and writing, and so forth. Now you tell me that is not so, that that is not a good judgment but rather it's the test.

Dr. **Rever**. No; I didn't say it was not. I said it may not be. The difference in a B in the high schools I graduated from in a southern little town in North Carolina and a B in the Baltimore Polytechnic High School is substantially different in terms of how people perceive the value of that B to be.

What is it worth in predicting how one will perform at Harvard, for example? The faculty at Harvard will say if you graduated from Baltimore Poly that B is worth considerably more than if you graduated from some small high school in North Carolina.

Mr. **Miller**. Unfortunately, if you lived in North Carolina, you couldn't go to Baltimore Poly.

Dr. **Rever**. That is true, but I still wanted to go to Harvard, and applied to demonstrate that it did not disadvantage me in what I learned, that I could compete successfully academically with those students who had been at Baltimore College.

Mr. **Miller**. The problem is Harvard's perception of your B.

Dr. **Rever**. Indeed, the test sponsors and agencies are offering an independent third party opportunity for graduates of all of our Nation's high schools to demonstrate what kinds of skills they have acquired, so if there is some selection that needs to be done it can be done fairly.

Even if selection is not to be done, they can make a judgment based on my high school records as well as my test performance to see what kinds of accommodations their curricula has to make to accommodate my particular development because of the high school I went to or whatever, the resources that were available to me to learn.

Mr. **Miller**. What about the test being overruled by Harvard, and you got into Harvard?

What about in the case where the tester told Harvard you shouldn't get in because you are from a little high school in North Carolina?

Dr. Perrone and others would suggest that there is not enough information to determine whether either determination is valid or not.

Dr. **Rever**. The amount of information that is available to determine whether the examination is valid?

Mr. **Miller**. To do an analysis of your operation's decision that this student, considering the high school, should be considered for Harvard or this student be considered for Harvard.

What about critics of the test who say that we don't know enough about how tests are evaluated, how the questions are put together, the rationale for the use of the questions, to know if that is a legitimate determination?

Dr. **Rever**. You have two parts to your question, if I may interpret that.

The one has to do with the release of test contents, and that is to say that the proponents of the bill have suggested that looking at the contents of the tests, one can then make judgments about whether or not it's valid, good, and does what it intends to do.

We submit to you that is an impirical question. If the intent of the examination is to predict first year performance in college for the purpose of selection, guidance, placement, whatever, we submit that these data are not hidden in fact. Consequently, we ask the question what do you mean by people don't know whether the examinations have any value?

We have substantial information, and we would be happy to provide the committee, which is inundated with research reports from every publisher, that in fact the tests do predict in a reasonable way how students will perform in college.

One has to make a selection decision who goes and who does not, and wants to provide the decisionmaker with the best possible information that is available, and that means both high school grades, college grades, recommendations of teachers, of friends, guidance counsellors, principals, and the like, so that that judgment can be made with the test and available possible information.

Mr. **Miller**. Dr. Perrone, do you agree with that?

Dr. **Perrone**. It's my sense if the tests were actually exposed and people could actually see precisely what it is that they are making judgments about in relation to students, that there would be a lot more doubts than there are at the moment, and I think there are considerable doubts already.

In terms of predicting, I think it's clear that ACT and ETS, College Board, will claim that their tests have some predictive correlation to the first year in college.

I submit that that correlation is not very high. It's relatively low. It is not as high as grades; grades are higher.

The argument then is that if you put both together, you are in better shape than if you have one or the other separately, and there has been some independent research that I do cite in the appended paper that I have given you, and I guess then you have to ask the question about equity.

Michael Wallach at Duke has done a good deal of research on what would happen if you used primarily the test score for purposes of admission, if you used the whole pool at Duke University, and if you used test scores exclusively, you end up with a population that even Duke at this point in time that would not support equity in any respect.

If you take grades, as Michael Wallach has done, grades and test scores, you get a configuration that is a little more diverse. If you take grades only you get a radically more diverse population, and then if you take what Michael Wallach calls achievement activities, what ACT and some of their research as talked about as nonintellective factors, you get a more diverse population.

Do the tests say very much about the kinds of predictions that one can make about a 4-year program or success in one's chosen career? ACT and ETS will not stand behind their tests in terms of anything beyond first year prediction, and there is a good deal to suggest the whole prediction breaks down when you go all the way to the baccalaureate level, and it suggests there is a good deal to suggest it breaks down when you try to do any correlation to success in any given field.

We need a much better definition of talent than comes through these particular tests.

I think we need much more equity than now exists.

Chairman **Perkins**. The gentleman from North Carolina.

Go ahead.

Mr. **Andrews**. It seems to me the institutions being referred to have done a thorough job, so far as I know, and far better than this committee is going to be able to do in the several hours that we might devote to these serious questions.

I think these institutions have within them, certainly Duke University, which happens to be in my district, and the others have people there far more qualified than we to determine which criteria should be considered in terms of the admissibility of the applicant.

Imperfect though it may be, I simply question the capacity of we who are on this committee or we who are in the Congress to in the short time available to us devise a better way of determining who should and should not be admitted than the institutions in question.

Second, I wonder if making available publicly the test, if that has, or as I am told and inclined to believe as its effect the relative destruction of that as one of the tools available; that is to say, that we will destroy the standardized testing as at least one criteria of the qualifications of the applicant, what are the alternatives that are available that are any better, and it seems to me the answer is rather obviously none.

Again, as imperfect as the standardized testing may be, if we in effect destroy it as one of the criteria, we resort to the other alternatives that have been allowed to or the only ones I know, that is, the achievement in the prior school or prior vehicle of education that the applicant might have been subjected to, and I guess in most cases we are saying high school, or the letters of recommendation from the various people, academic or otherwise, in behalf of the applicant and all of these are far more subject to rational criticism than is the standardized testing.

I guess my point is I am inclined to think that the standardized testing serves a valuable purpose which we should not inhibit by national legislation, and the extent to which it should be used as one of the methods of determining the ability of the applicant to achieve what he or she seeks to achieve can best be determined by those who represent the entity to which the applicant applies.

I am not in favor of approving either of the bills that are before the committee.

Chairman. **Perkins.** Thank you very much, Mr. Andrews.

Mr. Kogovsek?

Mr. **Kogovsek.** Thank you, Mr. Chairman.

The strong interest I have in the passage of H.R. 4949 is the insight that we can get, I think, as to what we are doing wrong in this country and in school distribution as far as low income and minority schools are concerned.

I would ask Dr. Rever and Dr. Snyder, specifically, first of all, while neither one of you actually came right out and said it, I seem to hear you imply, or got the impression that purveyors of tests have no responsibility to low-income or minority test takers.

If any group can attain the confidence that they need as far as test taking is concerned, I think it's these groups.

I think that taking a test is a traumatic experience for anybody, but it's so much more for a minority student who has been educated in a system that really has not brought him along as fast as he would have liked to have been brought along.

The openness that is involved in H.R. 4949 can give these students the confidence that they need, I think, in the future.

Let me ask you specifically if you do not think that it's the responsibility of purveyors of tests who take into consideration low income groups, I assume, you feel it's the responsibility of Congress and the educational people in this country, and I concur with that, but what is the responsibility of these people who put together tests that have to be taken by minority students when they are not prepared actually to take them?

In other words, what is the responsibility of a purveyor of a test and, second of all, is there any kind of a communication between those people who are making the tests right now and the school distribution that have a large population of low income or minority groups?

Dr. **Rever.** Indeed, testing agencies and test sponsors have been sensitive to the needs and the educational development of our minority and other low income groups in our society.

In fact, it's ironic that the one single index that points to their needs as a particular object of this Congress comes from the services we provide.

We do not intend, for example, to, given the kind of information and validity that the examinations have, we do not intend, it serves no purpose whatever to withdraw the examinations from the useful purposes for which they are put in helping minority youngsters diagnose what their educational status is and plan for the appropriate steps to be taken to be developed further.

We have been sensitive to the needs of minorities, both religious and racial and low income people, because our examinations indeed are reviewed by representatives of those people before they are even administered.

To suggest, for example, as Mr. Perrone has, that there is something easy about the linguistics of an examination, that makes it inherently biased is sheer folly and superficial.

Mr. **Kogovsek.** You are saying there is communication going on now between school districts, once a test is taken and results are sent back and some kind of communication that says your average students or whatever are performing thus and so, and we think this is why the minority groups and low income students are not doing well on the test.

Dr. **Rever.** We have a statement in here regarding that particular aspect.

We, as an agency, do not see ourselves in an evaluation position with regard to this secondary school system in this country. What we do is sample from the high school curriculum and try to make predictions about their educability at the postsecondary level.

On the basis of that, we trust the school system, the secondary school system, to make judgment of how important is that to them. If they judge that it is important, if they do in fact have a valid judgment about their students, they should take appropriate action and investigate why that seems to be the case.

Mr. **Kogovsek.** Dr. Snyder, you had indicated if H.R. 4949 does become law, you would like to see a year delay.

TEST VALIDATION

Dr. **Snyder.** So there would be adequate time for the promulgations, et cetera. Could I also respond to your first question?

The most direct way I can respond to it, this is not in the area of college entrance testing, but when we construct all of our major standardized achievement series, the ones that are used in elementary and secondary education, we routinely review all items with appropriate minority groups, but in addition to that, our particular company spends a rather considerable amount of money conducting simultaneous tryouts in various groups.

In addition to that, you raise the issue of uniform variety with the test, and to solve that problem, we have begun the practice many years ago of producing practice exercises for the examinations that would be administered to the students before they actually took the achievement tests.

We also provide essentially as a company service both pre- and postworkshops with teachers to counsel them in how test results can be most adquately used, specifically these days, in the class room.

The trend we are seeing is more standardized tests to be used in more diagnostic fashion than they were in the past.

Just one more thing: We have also introduced recently, several companies have the notion of a functional level testing procedure that can allow the adoption of a series of achievement tests to the actual level that the child is functioning at, so you get the most accurate measurement and can, therefore, use it in the most direct manner possible in the instructional stage.

Mr. **Kogovsek**. Are you far enough along to be able to tell those States who have bilingual bicultural programs, is it still too early to tell if these programs are working as far as your test results are concerned?

Dr. **Snyder**. We published on the bilingual evaluations an instrument that parallels one of our achievement series only about 1 year and 2 months ago. It's a bit early to tell. There may be some results coming out shortly, but I cannot say at this point in time.

Mr. **Kogovsek**. Thank you.

Dr. **Perrone**. May I respond?

Mr. **Andrews**. Yes.

Dr. **Perrone**. If you were to examine the power of testing in our society, I think one of the things you will find is that its power is related to the measure it controls and has access to all the information and holds it. As long as testing and clients control the information and it is not accessible for public scrutiny, testing is going to maintain a power that I think many people in our society are challenging without access to the necessary tools.

Comment was made it is very gross to talk about cultural and linguistic bias. I don't think it is. The fact is we don't have access to the tools. One can argue that minority groups quite regularly review items. But at the same time why is it that almost every major minority association in the United States has called in to question the whole testing enterprise and in fact gives support to bills like that in New York and I will suspect give support to the Weiss bill, et cetera, and to this particular Congress.

One of the questions they ask if you read the resolutions passed by NAACP, they don't absolutely assert that there is cultural and linguistic bias, they assert without more information we won't even know the extent to which it does exist. I think those are important issues.

Is there a problem? Take a look at the Parent-Teachers Association. I hope that this committee is sensitive to where the PTA has come from. If you were to have looked at the PTA 20 years ago, I think they would have been out in front of the testing enterprise without question and they were.

As it has gone through its own attempt to probe, as it has attempted to score from the testing industry what little bits of information it has been able to gather, as it has been reviewing the few items that it can get hold of, it has come to the conclusion that this is in fact the man without clothing, that there is something wrong, and no longer is the PTA the supportive body it was a few years ago.

My sense is if more people had an opportunity to review what is accessible, they in fact would see the power of that enterprise decline and come more into line with what is reasonable in our society as opposed to unreasonable in our society.

In terms of information, I might only add, I hate to use analogies like nuclear energy, but for how long did we believe that the nuclear energy industry had under control the safety issues relating to its employees and the public at large?

How long did it take to find out what impact there was in Utah from the 1950's testing? Would sunshine have been useful all along? I would suggest that sunshine would have been useful, and we might have taken even nuclear energy with a little more caution.

I think this particular body needs to do the same thing. The legislation does call for a marked change in a relationship of the Congress and the National Government to a private enterprise and to nonprofit institutions or agencies.

On the other hand, if you go to that fundamental issue that individuals who are to be judged by test results are entitled to know the basis of those judgments, it seems to me all the other concerns evaporate and ought to evaporate.

Mr. ANDREWS. Very good. The time of the gentleman from Colorado, I am told, has expired, and the Chair will now recognize the gentleman from Pennsylvania.

Mr. GOODLING. Thank you, Mr. Chairman.

I forgot to ask that the entire editorial be printed in our minutes of today's meeting.

Mr. ANDREWS. Without objection.

[The editorial referred to above follows:]

[From the Washington Post, Aug. 27, 1974]

SCORING THE ADMISSIONS TESTS

At first glance, it's a terribly appealing idea. A House subcommittee is now briskly proceeding with a bill giving students a legal right to see their college entrance examinations, with the questions and the corrected answers. For good measure, the bill would also establish broad federal supervision over all admissions testing. On a second and closer look, this legislation becomes less appealing.

Some of the support for it comes from people who simply think that students ought to be able to review their exams and see where they fell short. But some comes from people who want to change the nature of the test, and the whole admissions process, on grounds that these are tilted against the poor and the minorities. Rep. Ted Weiss (D-N.Y.), the author of the principal bill under consideration, says that he doesn't want to regulate the tests. But his bill clearly lays the foundation for a regulatory system.

Young Americans take these tests by the millions every year for admission to college or professional school. The most widely used, the College Board's Scholastic Aptitude Test, is given more than 20 times a year throughout the country and abroad. Each new edition picks up a good many questions from the previous one, to ensure that the scores on a test given on one date will be comparable to those of another.

But Mr. Weiss' bill would require the College Board, after each test is scored, to give the student both the questions and the answers. Since the questions would immediately be passed around to other students, they could hardly be used again. The cost of the tests would go up, and the reliability of the scores would probably go down. If you think that competitive admissions test are wrong in principle, fine. Certainly test scores aren't the only criterion for admissions. But to diminish their usefulness would force colleges to depend on the other and more subjective measures—not necessarily an advantage to the youngster who doesn't fit the usual pattern.

Mr. Weiss wants the scores reported to Congress by students' family income, race, sex and ethnic origin. What, precisely, do you suppose he has in mind? Its obvious that the children of educated middle-class parents tend to make higher scores than the children of poor and uneducated parents. But the increasing use of these tests has demonstrably been accompanied, even at the most rigorously selective colleges, by increased enrollments of children from disadvantaged families.

But not everbody likes the idea of an independent testing board run by the colleges that use it. The National Education Association, the country's largest teachers' organization, warmly supports the Weiss bill. Under it, the oversight of testing would reside in the new Department of Education, if Congress is unwise enough to create one. The strongest political influence within that new department would be the NEA. That's another reason for concluding that the Weiss bill has dangerous implications that go far beyond its author's stated intentions.

Mr. GOODLING. I think some people talked about excess profits on big campuses and I don't think this bill deals with that. If IRS is not certifying properly, that should be handled in court or by other legislation.

I don't know how familiar you people are or how much expertise you have in the business of looking over lesson plans, which of course, include teacher-made tests. I can only say in about 18 years of experience many of those tests frightened me. Of course, there is a quick way to do these things so you don't get too much paperwork involved. You can write essays and then the opinion of the teacher is the important thing, not the opinion of the student responding to the essay.

Would you suggest that somehow we proceed with legislation in relationship to teacher-made tests which are probably much more damaging than the tests we are now discussing. I believe we have pros and cons out there as I listened to the testimony.

TEACHER TRAINING IN TESTING

Dr. SNYDER. We recommended in other testimony that testing be added after development in the Higher Education Act of specific sections that are given in the testimony. The purpose of that recommendation was to authorize teacher centers to provide training to familiarize teachers with development in testing as well as curriculum testing and educational research.

Mr. GOODLING. In other words, you are saying all those courses we had to take through college and gradute school really weren't very helpful.

Dr. SNYDER. I am not sure whether the courses weren't helpful or whether they disappeared from a lot of the teaching curricula.

Mr. GOODLING. Anyone else want to respond to that?

Dr. POTTINGER. I would say, Congressman, that if tests a teacher made up as the gatekeeping device was the one-shot decision that affected whether that child had an opportunity for higher education or work opportunity, they should be scrutinized.

The fact is that teacher-made tests that are used routinely are not that influential or don't have the impact.

Mr. GOODLING. Wait a minute. I heard a lot of discussion about getting away from any kind of standardized testing. Wouldn't grades be the sole thing and that would be extremely important in relationship to someone's admission into college if you go that route?

Dr. POTTINGER. I don't think that the issue is that we should have tests. I think the issue is which tests should be scrutinized, evaluated, and understood.

Mr. GOODLING. Then, of course, it is very important who does the scrutinizing, who has which axe to grind.

Second question: At the present time we will have legislation somewhat similar to this in operation in New York. We have some kind of legislation in operation in California. We have the National Science Academy presently studying this whole issue.

Shouldn't we wait to evaluate their results, and use what they find out from their experience before we as a Federal Government jump on this bandwagon?

Again, I think you are divided in camps down there.

Dr. SNYDER. My response is that the issues are complicated enough and since there has not been a demonstration of any sort of emergency condition, it would be best to wait until all the evidence is in. The bills themselves talk about analyzing data, seeing how something goes with respect to State statutes, seeing how a review goes. The National Academy of Sciences would certainly be data that would be helpful in the decision process.

Dr. REVER. Indeed, the experience we have in New York and perhaps California, although the legislation, as we pointed out in our testimony, is quite different from what has been introduced here, would be appropriate.

We have in fact listened over the previous 2 days of testimony given today in which the beneficial claims for this kind of legislation are highly speculative. In fact, the proponents of the legislation say to us test agencies who are opposed to the legislation, I don't believe you either. You have no data on which to base your claims. We focus on such trivial matters as 10 or 15 cents more per test.

We don't know in fact what the effects of the legislation are going to be on the disadvantaged student, for example. I could argue I think in a fairly convincing way to most reasonable people that our educational system is not fair to all students. We make that point in our prepared testimony.

The ones who are hurt the most, as an index of high school grades show, are related to socioeconomic status and I have a table in my testimony that relates to that.

Why would we assume that by making the examination public we will automatically help and be of great benefit to the disadvantaged in this country? The legislation that this subcommittee has led the Congress through to help the disadvantaged in this Nation is often disappointing in the evaluation of it. Even for those Federal programs that are designed to help the disadvantaged, we can't always demonstrate a positive response, so why would we assume that all of a sudden by making all manner of data available, particularly test contents, the disadvantaged indeed are going to be helped, particularly the students?

The proponents will argue, of course, that revealing the contents of the examination will throw open the tests to all kinds of research for immediate improvement and that will not be a problem. We are not sure that is the case.

Mrs. SPARLING. I would like to respond that there is legislation in some States. As a parent and expressing the concern of parents in the other States, we would like to have this particular bill because we can have the information that allows us to make the kinds of decisions about our children's education that we need to be making.

Mr. GOODLING. Would you not rather have the kind of information that has been proven rather than the kind of information that may be totally incorrect? Let me give you an illustration. We did away with the teaching of phonics, for instance. That was strictly motivated by the Federal Government because of the money it provided for that kind of experimentation.

It would have been better done as a test kind of thing, I suppose, on a few rather than spread all over the Nation. That is one of the concerns I have here. Do you not want legislation that you can be sure of?

Mrs. SPARLING. In this particular bill it would give us the opportunity to have the kind of information that would indeed help us to make these decisions in H.R. 4949.

STATE AND FEDERAL EXPERIENCE

Mr. MILLENSON. With respect to your question, as we pointed out in our testimony the legislation before this committee, both bills go far beyond the New York State law so that even the experience in the New York State law should be relied on.

No. 2, if this legislation is desirable, then also in addition to the State experience I would think you would want experience from the Federal Government. The Federal Government gives tests. You are skipping the Federal Government and going right on to imposing this elsewhere on States and localities.

I might add, not only the Federal Government but as was pointed out here before, members of this committee who rely upon tests themselves for appointments to the military academies, why don't they apply this first, see how it works with them and maybe we can go on?

Mr. GOODLING. I won't have time for a response to what you were just talking about but one question that came up during the discussion, concerns this business of eliminating standardized tests. If we get to the point where they are no longer a part of admissions—and I certainly don't think they should be the only criterion—but if we go away from that, will not those schools such as Duke University, for example, have a tendency to be very selective from which schools they choose their applicants? There are many ways to handle this matter. As I said, I will not have a chance to get your response. You may want to respond in writing.

The second thing was just touched on. I have written to the chairman just recently and said:

> It seems that before this committee considers marking up any legislation designed to regulate standardized testing which will impinge solely on the private sector, it

ought to consider the impact of standardized tests administered by the U.S. Office of Personnel Management, the Armed Services, and other Federal agencies.

If, as the sponsors of the testing legislation contend, it is a matter of public interest to apply certain testing standards to the private sector, it is even more important that this Committee investigate the ramifications of applying such standards to the Federal Government.

I would say, again, you will have to respond in writing. I am running out of time. I am wondering whether that shouldn't be our main avenue because we surely do affect an awful lot of lives with the kind of testing that takes place in the private and public sectors. Maybe we should start at home. I am not sure about that.

Just one other thing. Again, you can respond in writing. I have a concern about combining these two pieces of legislation. In your reading of these bills, some people would suggest that unions then would have to submit such information as copies of contracts, revenues from testing programs, and testing overhead spending.

I can see if this becomes an issue of that nature, knowing this committee and the makeup of the Congress, we are going to have a real problem if there is some impact on the unions.

Dr. **Perrone**. It seems to me the legislation before us is fairly limited. It doesn't deal with high school tests given by teachers, doesn't deal with State competency standards or tests for promotion from one grade to another.

Mr. **Goodling**. But there is a lot of talk about combining two bills.

Dr. **Perrone**. Let's take both of them. Both of them are fairly limited in that sense. They are not addressing all the standardized tests given to elementary and secondary schoolchildren in the Nation. The bill is fairly focused on postsecondary testing where at least large numbers of people perceive a serious problem to exist.

One can take on that question without at the same time thinking that you must first take on the entire world. It is prescribed. It is prescribed, I think, for a number of legitimate reasons.

Mr. **Andrews**. Thank you both.

Your time, as you have indicated, has expired.

The Chair will recognize the gentleman from New York.

Mr. **Weiss**. Thank you very much, Mr. Chairman.

Just to reemphasize the point that Dr. Perrone had made, this legislation in no way applies to occupational areas. It is simply limited to admission to postsecondary institutions. Indeed, Dr. Snyder, in your testimony I think you acknowledge the fact that while only H.R. 3564 applies to specific tests produced that your concern as to H.R. 4949 is that some of its provisions might well be incorporated into final legislation.

You have no concern about H.R. 4949 applying to any occupational tests, do you?

Dr. **Snyder**. My concern is if H.R. 4949 provisions are included in any bill for occupational testing, then my comments would apply.

Mr. **Weiss**. You do agree right now it does not address occupational testing?

Dr. **Snyder**. That is correct.

Mr. **Weiss**. As far as Mr. Goodling's statement, I just want to reassure him and everyone else I have not heard of any intention of joining the two pieces of legislation, so regardless of what may or may not be in somebody else's mind, it certainly is not in my mind.

Dr. Rever, let me ask you a question. Can you tell us roughly how many examinations for admission to postsecondary institutions are given on an annual basis?

Dr. **Rever**. Mr. Weiss, that is difficult to estimate for several reasons. One, we know that the college entrance examination board, I believe in their annual reports, said for admissions purposes it was something like a million and a half. For the assessment it is about 900,000 each year. In addition to that, you have the intrastate college admissions testing programs which I referenced in my earlier remarks which must offer services maybe to several hunderds of thousands of students; but I have no figures.

Mr. **Weiss**. It runs into the millions, I would assume.

Dr. **Perrone**. Yes. I don't have a precise figure for you.

Mr. **Weiss**. I think Dr. Rever had indicated there was a concern about the constitutional issue in the situation because, after all, what we would be doing would be addressing strictly intrastate situations between institutions and the students within a particular State. Is that your position?

Dr. **Rever**. No. We say there are essentially two parts to that. One has to do with the fact that it is simply a State's right political issue.

Mr. **Weiss**. What do you do with the situation where, for example, in New York right now the medical people have taken the position that because of the existence of increasing New York applications they will not give tests in New York upon applying to medical schools.

Would Federal legislation be helpful in that instance?

Dr. **Rever**. Not likely. I cannot speak for the Association of Medical Schools.

Mr. **Weiss**. If in fact the Federal law requires disclosure, no matter where that testing organization gave the test, would there then be any advantage to MCAT saying we will move to New Jersey and give our tests there?

Dr. **Rever**. I don't believe AAMC has made that statement. What they said was if your bill were to be passed they would indeed have to change the nature of their examination as it is now offered to be able to sustain service.

Mr. **Weiss**. No. MCAT people testifying before us, Dr. Cooper, specifically stated that if in fact the legislation passed in New York were to go into effect and be implemented and not repealed, they would in fact move out—they would not give those tests in New York.

Dr. **Rever**. Oh, yes, that is true, but you have asked me what would be the effect of Federal legislation on AAMC. What I am saying to you is Dr. Cooper has responded to that question and he has indicated that the nature of their examination, their new MCAT would have to change to be responsive to this legislation. They object to that because they feel they are not confident in the nature of that examination. They would have to revert to the former type of examination.

Mr. **Weiss**. How long has the new MCAT examination been in effect?

Dr. **Rever**. Two years.

Mr. Weiss. On the basis of that you are able to say it is a better examination.

Dr. Rever. I have not done an evaluation of that. That is AAMC's responsibility and I cannot comment on that.

FURTHER LEGISLATION ON TESTING

Mr. Millenson. Mr. Weiss, I would like to respond further to the question you posed to Dr. Snyder and perhaps we can get some clarification of our concern that your bill could possibly be applied to more than just college entry.

In your opening statement at the beginning of this series of hearings, you said, "The testing industry," which comprises the companies we represent, "despite its influence over a vital aspect of individual lives, operates in a most unaccountable fashion. Like utilities, testing corporations perform a public function and they say you are exempt from public scrutiny."

Then on page 3 of your opening statement you refer to the fact that 50 million young Americans will take up to three standardized tests during their academic careers. That refers not only to college entry tests but the elementary and secondary school tests that our companies provide.

Then you talk about at least $200 million in annual revenues.

Mr. Weiss. To allay your concern, on page 13 of the bill, at line 6, the term "standardized test" or "test" means: (1) Any test that is used, or is required, for the process of selection for admission to postsecondary educational institutions or their programs; or (2) any test used for preliminary preparation for any test that is used, or is required, for the process of selection for admission to postsecondary educational institutions or their programs.

Mr. Millenson. I am aware of that but also on page 3 of your statement you refer to companies which are not members of our association which give tests which are not college admittance tests. So the fact you refer specifically to them raises a very natural question in our minds that perhaps there might be some intent if not now and specifically in your bill, which I have read in detail, at some later time to apply these restrictions not only to college entry tests but to other standardized tests, especially as we see some of these statements that have been made, for example, in the New York Times, August 15. "A passage of the law is a triumph for Ralph Nader and his associates who have been campaigning against standardized tests in general."

Mr. Weiss. You are satisfied now that as far as the information before you, H.R. 4949 applies only to admissions.

Mr. Millenson. I am absolutely satisfied of that. I am not satisfied as to what further intent might be with respect to this legislation or how it might be expanded or what the basis might be for future action.

Mr. Weiss. What you are doing is sounding an early alarm.

Mr. Millenson. I am indeed doing that, sir.

Mr. Weiss. Dr. Rever, can you tell me what percentage of any particular test your organization gives is new?

Dr. Rever. In our particular test there is—let me explain how we do this.

In our method of equating examinations, which is what I think you are asking——

Mr. **Weiss**. No; I am asking really for those questions which are used in developing the raw score on the basis of which the scores are given to the students as to how they did on the test. Not new questions. There are equated questions, there are experimental questions. I am not concerned about those. I want to know those questions that are used on the test which in fact are new in any particular test that are used with a score given to the student.

Dr. **Rever**. May I continue on?

Mr. **Weiss**. Since we are limited in time, try to give me a percentage if you can.

Dr. **Rever**. A hundred percent for four or five test dates.

Mr. **Weiss**. A hunderd percent of the test questions are new.

Dr. **Rever**. For four of the five test dates.

Mr. **Weiss**. In any particular year?

Dr. **Rever**. Yes. That is why I wanted to explain. The way in which——

Mr. **Weiss**. So even without the disclosure in four out of five instances you provide 100 percent new questions in each test.

Dr. **Rever**. Virtually.

Mr. **Weiss**. There would be——

Mr. **Andrews**. He says he wants to explain that you are misinterpreting his answer.

Mr. **Weiss**. I hope I am not.

Dr. **Rever**. The question is related to the manner in which we equate forms. We do equate all of our forms at a certain test date during the course of a year.

The evaluation occurs well before the examination is entered, so roughly there would be some equivalent to cross the examination in the duplicative items, but the percentage is relatively small.

Under revisions that would be required under your law, we would estimate that approximately 20 percent of the items would be repeated items for that purpose and 80 percent would be new.

Mr. **Weiss**. Right; so that it would be 80 percent whether we had this disclosure requirement or we didn't have it.

Let me put it by way of question. On that basis, how much does it cost additionally to develop that 20 percent of new questions?

Dr. **Rever**. Our position is that the additional cost is marginally spread over the millions of students we are talking about, but we did make a point that one must be careful in looking at the law about the impact it has on small value tests, for example, braille tests.

The cost from our perspective is not the central issue of the legislation; and this legislation, its desirability or undesirability, should not be based on the cost of developing different forms of the examination. That is a misdirected argument.

Mr. **Weiss**. You are aware of the fact that H.R. 4949 provides an exception to those tests which are given to 5,000 fewer parties?

Dr. **Rever**. We do have more than that in terms of special administration, but different versions, for example, a braille version, and so on, and examinations are administered essentially 365 days a year on the shifts and the like, and the numbers and interpretations of the intent of your law does cause us a considerable amount of concern, because it's all called the same test.

Mr. WEISS. One final area of questioning.

Do you have any information to give us as to the margin of error that is built into any of these academic admission tests?

Dr. PERRONE. The margin of error tends to be fairly large for most tests that I am familiar with.

For example, most of the standardized tests that exist, the argument is that in two-thirds of the cases you can judge that the standard of error will fall on either side of the main by anywhere from 3 to 4 raw score items.

On the SAT I suspect that it would range, a score of 450 would probably be very much like a score of 550.

That is the range in which the standard of error would accrue. It tends to be fairly large.

I think that it has been a mistake for test companies to give a single score. I think that has been part of the misleading aspect of the scoring system.

Mr. WEISS. So that if a school which uses one of these tests decides hypothetically that it will consider for admission those students that score 600 or above, and a student comes in with 520, that that student may in fact be excluded on the basis of the margin of error which exists in the test itself and not the performance of the student?

Dr. PERRONE. That is correct. That student may very well be excluded by some institutions for that reason.

Mr. WEISS. Dr. Pottinger, do you care to comment on that?

Dr. POTTINGER. We once asked some members of the Educational Testing Service why they didn't present scores to clients, to teachers, using an interval score; that is, saying that the probability is some high percent that a student's score falls within a certain range to try to break down this, the aura that somehow a 475 is better than a 450, and in fact it may not be.

Yet schools may make decisions on the basis of that difference.

I think that is a critical problem in the test. The response by the Educational Testing Service was that they provided scores and some of their tests in interval ranges and the client said, "We don't understand this. It's unclear. Give us a single score;" so the testing companies are not in the business of providing information, teaching people better ways to use the tests.

They are in the business of providing a product that the client asks for, but the fact that the client asks for something simplistic or that they can make decisions that in fact may not be right seems to be of no responsibility to the test company, so there is no incentive for them to provide that information, even though they recognize that that would be a preferable way of presenting the scores.

Mr. WEISS. Dr. Rever?

Dr. REVER. If I can respond to that?

In defense of either Mr. Millenson or Dr. Ferguson, I don't know which, this is Mr. Millenson, and this is Dr. Ferguson [indicating].

Mr. WEISS. My apologies to either or both of you.

Dr. REVER. The issue is related to what kinds of scores are reported; and, in fact, I think, Mr. Weiss, if you will look at the kinds of scores to which Dr. Pottinger was referring, those that may be used for guidance or selection in postsecondary institutions, ample information is provided, and the data are reported in terms of intervals.

In fact, we report the data typically in much better terminology, and that has to do with probabilities of success which are a lot more meaningful to most students and admissions officers in terms of how one deals with those data.

Mr. **Weiss**. What does the student receive back after he takes the test?

Dr. **Rever**. I will be glad to provide you many copies.

Mr. **Weiss**. As far as the score?

Dr. **Rever**. An estimate for up to six schools that he can choose of his choice, what his chances of success are in terms of academic grades, what the local norms are for the examination, how do his high school grades and scores compare to his future competitors within that institutions, his scores and the national percentile ranges for his school, how do they compare to the entire population of students who have taken the examination for the previous 3 years.

In addition to that, there is something like probably 190 different, separate, distinct pieces of information that are reported to the student including everything from what the opening date of the school is to what it costs to attend.

Mr. **Weiss**. Either on your test or on the SAT, does the student receive back the numerical score?

Dr. **Rever**. Indeed, they do get numerical scores for each of the subtests that are reported in an average.

Mr. **Weiss**. By number, 475, 525?

The student is not advised that in fact this is the range?

Dr. **Rever**. Indeed they are in the handbook for students, and I will be glad to provide you one.

Mr. **Weiss**. Before or after?

Dr. **Rever**. As they receive their test scores they are provided interpretative data.

Mr. **Weiss**. I am sorry. Mr. Millenson, we have just had occasion to check with the Naval Academy in response to a comment, and you will be pleased to know the SAT and the ACT, and the other academies do as well and, therefore, the likelihood would be that they——

Mr. **Millenson**. I am referring also, of course, to Members of Congress, the procedures which they use, too.

Mr. **Weiss**. Those are occupational tests.

Mr. **Millenson**. This bill would cover the service academies?

Mr. **Weiss**. SATs.

Mr. **Andrews**. Are there other members of the committee who have questions?

Mr. **Erdahl**. One question, and maybe I can address this to a specific member of the panel.

Dr. Pottinger, maybe I know the answer, and I am not a lawyer, but I know lawyers ask questions like this.

Do institutions have to use these tests? Couldn't they use the grades, their own entrance examinations? Do any institutions have to use these standardized tests?

Dr. **Pottinger**. Well, when you say do they have to, I assume you are saying isn't it a voluntary process. Some institutions choose not to use them.

Dr. **Rever**. If there is a better system, they are entitled to use that.

Dr. **Pottinger**. If they have the information about better systems and if a better system exists, if you can develop them, if the information is clear, of course, they would be free to do that.

The problem is that information is not available. We don't know enough about the tests, and in any specific example, the fact is, to get the information, you need to make reasoned decisions about the test which often means you have to go to court and spend years in court to get the data you need.

The critical issue, you raised the question earlier that relates to this, and that is important. Who is getting hurt and who is not by the tests?

That is probably our major concern.

To use a couple of concrete examples, it might help.

The Foreign Service, I realize this is an occupational exam, but the example gives us a multiple choice test, and for a number of years the Foreign Service because of that test yielded a highly white middle class, often ivy league corps of people which was acceptable, but in the sixties that notion came under question.

They waived the requirement of the Foreign Service test, because minorities and others who did not do well on the test needed to be given access to those jobs.

They discovered some years later that those people who did not take the tests, some of them became their star Foreign Service officers.

They came to us in another firm I worked for at that time and said: "Why is this happening? Why do we have people who do badly on the test doing well on the job, and are we in fact discriminating?"

The study that we did for the Foreign Service said that the problem is the test items compare and contrast with a Mozart and Shubert scenario and simply did not relate to the job of doing Foreign Service work.

You don't know what those items are about, and what the characteristics and how it relates to outcomes, if you don't have information.

Second, I think that the people, the point I am really trying to make is those that are being protected often are the people who have learned to take tests or learned to do well on specific kinds of contents that have little relationship to the more important consequences of education, work and living, and I think the notion that grades are SES-oriented, correlated with socioeconomic status is equally true of multiple-choice tests.

You read David McClelland's article in 1972 called "Testing for Competence Rather than Intelligence." There are other examples in the human services, particularly mental health where people use inner-personal skills effectively, but we test them for esoteric knowledge analogous or sort of irrelevant kinds of esoteric kinds of information.

Mr. **Erdahl**. What are the alternatives to a standardized test that are workable? Somebody brought up the essay-type tests. Anybody who claims that the standardized tests are the only thing you want to look at, would be in error if they limit their consideration to this.

It should be grades, personal contact or interview, but what are you suggesting?

Dr. **Pottinger**. I would ask the universities and colleges who don't use the test what their mechanisms are. I can tell you from our own research, the preferable alternatives are measures where people have to demonstrate the very skills that they are being asked to use in their work and life responsiblities.

It's interesting, I think, that the reason multiple-choice tests——

Mr. **Erdahl**. I hope you don't carry that analysis to a brain surgeon, where you have to demonstrate before you get into medical school.

Dr. **Pottinger**. I would argue the other way recognizing correct items on a multiple-choice test which is the way we now certify brain surgeons is probably even less adequate.

The reason multiple-choice tests do correlate with the first year of school is because the very way we operate in education, especially in the early years is multiple-choice tests.

Once we get into the real world, we are not confronted with information in a multiple-choice format where we have to recognize the correct alternatives. We are confronted with a number of domains of behavioral and affective qualities that cannot be measured by multiple-choice tests.

We are saying we will allow people the opportunity for work as professionals or whatever the schools are admitting them to, if they can demonstrate that they can get through that first year of school. Forget about the rest of their qualities that will be required of them in their later years of school and on the job.

Mr. **Erdahl**. We are getting into a brandnew concept here that everybody then should be entitled to enter a college or university. You are saying that in a sense if you don't make it the first year you don't make it, but at least you have that chance to go for 1 year.

That adds a new dimension costwise.

Dr. **Pottinger**. I am only focusing on the inadequacy of this particular method for gatekeeping. There are beginning to emerge work samples, what we call operant measures rather than respondent measures where we ask people to demonstrate something.

One reason these methodologies are not being utilized is because the inadequacies of the current system are not being exposed.

Mr. **Erdahl**. I would hope institutions would use a broad spectrum of different criteria in admitting students.

I spent some time at the Mayo Clinic over the August recess, and they have a rather exclusive medical school there, because they have several thousand applicants for 40 members in the class.

There the grade-point average of the people that apply is probably close to 4 points.

I talked to one of the people there, who said they use these various things, personal interviews, and for the out-of-State students they allow five new members each year. They have 1,400 applicants, any one of which maybe could make it.

They go by standardized tests, school grades, interviews, and try to broaden out a class as far as different ethnic and geographic backgrounds.

I would hope, and I don't think this is the gentleman from New York's intent at all, that they don't exclude the possibility of using the standardized test as one of the tools.

I think also, Mr. Perrone, you are eager to make a response.

Dr. **Pottinger**. I would agree with that. The example you have given is an excellent one, and we would concur. We are not saying we should throw out all standardized tests, but saying we should open them up for analysis so we can know when they are good and when they are not.

I can cite examples of law school deans of admissions who have said: "I don't have time to do that, and anyway my alumni association, if they know we were taking people who make below 650, would be all over my back."

That is the other side of the problem.

Mr. **Andrews**. When you say we should open them up so that we can know when they are good and when they are not, who is "we"?

Dr. **Pottinger**. To begin with, it's the students who take them.

Its researchers want to be better informed about the effects, for instance, the discriminatory effects or the relationship between those measure and the socioeconomic status.

We are put into a double bind, and when are the proponents of this bill going to show some data that these tests are discriminatory. How can we show the data when we can't have access to it?

We don't want to be frustrated by the need to go to the courts and pay lots of money. Most of us can't afford the thousands of dollars that is required to go to the courts to get this information.

Dr. **Rever**. May I respond to that, sir?

Dr. Pottinger knows full well that the issue of hiding data relating to bias and the like is certainly not the case with testing agencies that I deal with.

We have never, ever turned down a responsible researcher from taking a look at the contents of our examination or any of the data that we have in our files.

In fact, we provided, upon request, people with significant data sources which they can make judgments about all kinds of things to which testing is related, some of which has to do with learning, by the way, but that is part of it.

The data that we hide is hard for us to understand, you see, because when you talk about the issue of is the examination biased against certain classes of people in our society, in fact the test agencies have been the ones who have been the leaders in analyzing the effects of their examinations and the use of their examinations on those very people.

What we do in fact find is that the examinations, when used for selection, often work in the favor of the disadvantaged rather than vice versa, and so to claim that there is untold and unknown kinds of disastrous effects on society through the use of examinations is indeed well researched.

What the proponents really want to talk about is the contents of the examination.

I cannot tell whether the examination is inherently biased against certain parts of our society unless I see the contents of the examination.

Well, we have faculty members every year by the scores who look at the examinations, and admissions officers look at the examinations as a part of choosing which examinations are they going to require for their students.

To assume that they do not know the contents, and it has not been studied by academic committees who are in charge of admissions, is really missing the point.

Mr. **Andrews**. Is there anything contractual or otherwise between the testing, for a lack of better terminology, company, whoever makes the test and the schools, say, a subscribing school?

We will take Rochester, whoever it is. Is there anything that precludes the people at Rochester, the members of the admissions committee, the dean of admissions, or whatever, doing with that test whatever he or she wishes?

If they want to go to AP or UPI or something, and give copies of the test to anyone who might ask them, is there anything that precludes them from doing that?

What are the exponents of secrecy here?

Dr. **Rever**. In general there are copyright laws; they could not reproduce the examination which falls under copyright law. In most of the documents, all of the examinations are in fact copyrighted by a test sponsor.

Mr. **Andrews**. What does that mean exactly? That means they couldn't reproduce it and sell it?

Dr. **Rever**. I believe it also means distribution of any kind. They could not reproduce it and distribute it at any time.

Mr. **Andrews**. Where the student who feels he or she might have been discriminated against, could that student go to the appropriate office of the school and say: "I want to see a copy of the test I took and on which I received a low grade. I know I saw it the day I took it, but I brought my parents with me and we want to see the test again."

Is there anything that would preclude the appropriate official at that institution to say: "Here is a copy of the test."

Dr. **Rever**. Often he has an equivalent form that he could share with the student. There is nothing that precludes them from doing that in our policies. In fact, students upon request cannot see a copy of the examination unsupervised and the like, but we have had attorneys and students who show up for particular examinations and we sit down with them both and go through the examination.

Mr. **Andrews**. It seems that the bill would cause the test to be made public, and I guess my question is, are they now public?

Dr. **Rever**. No, sir; they are not to the students. The publication of the tests themselves are under the control of the test agencies and the test sponsors, and they are not public record at all.

We control who sees them and who does not, the purpose of which is not to hide the examination.

Mr. Andrews. When you say we, you are talking of the testing company?

I don't think the student probably, or the applicant who is not now a student and is disgruntled by reason of the fact he or she wasn't admitted to be a student, that is the person we are trying to assist; is that right?

Dr. Rever. Right.

Mr. Andrews. You say that applicant does not have access to the test or may or may not have from the company that prepared the test; but the student, as I understand it, has no relationship with the company preparing the test?

The applicant has some sort of relationship to the institution to which he or she applied, and that institution, not the testing company, which has denied the applicant admission, is that right?

Dr. Rever. That is correct.

Mr. Andrews. Is there anything to prevent the applicant going to the institution and getting such information he or she might want as to why that institution failed to admit him?

Dr. Rever. Indeed they may.

Mr. Andrews. Would that not relate to the institution's right? The institution is not precluded from giving that information to the student?

Dr. Rever. That is correct, as far as I know.

Dr. Perrone. I have a sense that Mr. Andrews has really moved very close to the fundamental issue in the bill.

Ought the person who takes the test have an opportunity to see what it is that he or she is being judged by?

The precluding of that individual from seeing the examination, seeing the answers, seeing their own answers in relation to it is the fundamental issue in this bill.

Mr. Andrews. I don't know this, and I no doubt should, is this bill directed at the testing company requiring the testing company to make the test public, or is it directed to the institution, for instance, Duke University, who denied admission to the applicant?

Dr. Perrone. In the case, for example, the law school admissions council is technically responsible for the administration of this law school aptitude test. That is prepared on their behalf by the educational testing service.

I might say they do have a process for review. The process is long and complicated.

A student who might want access to the LSAT would be in a process for months and months, and then might see the exam under conditions hardly conducive to a serious review and long past the point of it doing that student a lot of good.

Mr. Andrews. Does this bill require the testing company to make public the test, or does it require the appropriate officials at Duke University to make public the test?

Dr. Perrone. The officials at Duke University do not have the test.

Mr. Andrews. Well, let's pick an institution that does not.

Dr. Perrone. ACT has the test; ATS has the test, the law school admissions council may have the test, the American Medical College may have the test.

Mr. **Andrews**. Well, my question is to whom is the directive in this bill directed?

Mr. **Weiss**. The test agency, and the test agency may indeed be the testing company as in the instance of ACT or ETS. It could be the law school admissions council, the agency, but it's not the institution, and Dr. Rever may have misunderstood perhaps some of the questions, because I think in the course of the dialog, gave responses which may be ambiguous, and perhaps he can try to clarify that.

OK, again, is it not true that it's the testing agency which has the ultimate proprietary right in the examination which is given to the student?

Dr. **Rever**. Yes.

Mr. **Weiss**. After the student completes that examination, the examination goes back to the testing agency itself.

Dr. **Rever**. Yes.

Mr. **Weiss**. It's the testing agency which then supplies what to the institutions who are the using institutions?

Dr. **Rever**. Score reports and interpretive manuals.

Mr. **Weiss**. Right, but not the individual applicant's test and results.

Dr. **Rever**. That is correct.

Mr. **Weiss**. The applicant can in no way go to the institution because at the time he takes the test he may have indicated a half dozen institutions to which he would like to have the test scores sent.

Dr. **Rever**. Yes.

Mr. **Weiss**. This bill is in fact necessary if we were in fact to have the kind of accessibility that Mr. Andrews indicated the student or student's parents would like to have as to how the student did on a particular examination, and what he did in relation to a particular question; is that right?

Dr. **Rever**. That is correct.

Dr. **Pottinger**. I want to make a simple elaboration, because I think, Mr. Chairman, you have fingered the issue as he says, and if we look at an item on a test we were involved in a case where we were finally allowed to see the items.

An expert in the industry that the test was developed for saw an item on the test, and I am using one item as an example that could represent many.

This item has nothing to do with what this outcome is to be predicting.

In my opinion it just is unrelated. It happened to be a case where the item related to information that had been out of date for 25 years.

The point is that so-called objective tests are at the base very subjective. Someone has to decide what are the items. What are the questions we are going to ask?

Who are those people? How are they selected, and what do they ask?

If you simply take the test developer's word for it, and you may get the names and even the addresses perhaps of the people who made up the items, but what is it? What are their values? What do they believe is important?

The point I am making is that there are times when some of us feel that their subjective opinions about what goes into the contents of those tests is inappropriate.

We have no way to challenge it, unless we go to court and go through the mechanisms that we have heard about.

We simply cannot make a straightforward challenge to the subjectivity that is inherent in these tests.

What is objective about them is not the content, but the way they are given.

Mr. **Andrews**. May I ask this question? If the person or persons who made the test that you now perhaps question, assuming that you or somebody comes to the conclusion that certain questions were not relevant or the test was in some way not fair, what do you prescribe be done, that you designate who make the next set of tests or who is to make the designation of which subjectivity the applicant is going to be subjected to?

Dr. **Pottinger**. I think there is an intermediate step. Before we can answer the question of what is a fair process of arriving at the decision of what content should be on the test, we have to know what the content is. Once there is a body of professionals or other experts who say we think this is contrary to the best interests of students going into this school or into this profession, then we can have an open debate and the process may in fact come down in favor of the test as it stands.

That would be OK if in fact the consensus or some due process led to that conclusion. We simply don't have a process that enables us to have that debate or to challenge the people.

The test developers often are the people responsible for picking those who write the item. Therefore, they in a sense are responsible for the content of the test, indirectly or sometimes directly.

Again, there is no way to challenge that process. It is a tremendous power of an industry. If we were building automobiles the way we are building tests and we only allowed a certain few experts to design the parts and kept them secret, we would be in a real mess.

We do not have innovation. We would not have very good products probably. Yet the industry and some experts would say this is the best car we can produce and the experts agree with it.

The fact is we have to open those decisions up for debate and due process.

Dr. **Perrone**. The specter that is before us is after possibility and after open debate these examinations get strengthened, they may get strengthened to the point where they are so credible they may even have more power in the American society than they have already. I can see that potential.

There is the other potential that the alternatives will be examined more seriously. It is very difficult for an institution, a college or university, especially those that are not in the top 10 or 15, to go it alone in alternative directions when the wave is all somewhere else.

Bowdoin could easily discard the SAT as it did by saying it doesn't relate to anything that is important to us. Hampshire College could discard the SAT. Harvard University and College can say, as Dean Whittler has said, we no longer use the SAT to the degree that we did 15 or 20 years ago. It doesn't meet our purposes anymore.

But there are 2,000 other institutions in the United States that in fact through their own accrediting associations have to show some kind of quality mechanism and they refer to this process as their way of doing that.

I think historically we have examples of procedures that are different than those that we tend to be trapped into at the moment. The 8-year study in the late thirties and early forties provides, I think, a very good example of alternative procedures for showing colleges and universities what people in fact can do.

They were much more operant in David McClelland's terms than the tests we now use. I am convinced that those alternatives, though known about, are not used to any great extent today because of the power of that process which is already in place.

If we are going to have the serious debate that I think educationally and politically and culturally we need about these powerful forces in people's lives, then we have to expose everything that exists. To leave any of it hidden is not in the best interests of the society in any form.

I wanted to respond to Mr. Goodling and Mr. Erdahl in part, a question that aren't we going to encourage a coaching industry.

I might only suggest that Mr. Kaplan, who has the largest of that bit of industry, has not shown up anywhere that I know of to argue in favor of any of these bills. In fact, it is fairly clear that the bill that is before you might in fact do damage to that kind of enterprise by making it possible for the public sector to carry out more effectively what Mr. Kaplan seems to have a particular hold on at the moment.

I think in fact we will see a different process than more of that kind of industry.

One last comment: The one thing that has not been said, and it stays with me because I work with large numbers of people who have not typically gone to graduate schools or professional.

And one of the things that has been very disconcerting is trying to encourage individuals whose experience with these kinds of tests have been so negative that while qualified in every respect, they won't take the GRE, they won't take the LSAT, they won't take the MCAT, because they absolutely don't want to go through what they perceive to be a humiliating process that they have not had any kind of success with in the past.

Yet their qualities would make fine attorneys and fine doctors and yet they have not even made the step.

Mr. ANDREWS. On what do you base that judgment—you say they are well qualified—by your subjective standards?

Dr. PERRONE. I am going to suggest to you that their qualifications don't in and of themselves permit that entry. There are a number of schools that demand that test be taken and there are a number of people whose experiences have been so negative that they don't take them.

I am suggesting that is the power of that particular piece. It is a power that I think is far greater than we ought to tolerate.

To the degree that we open the process up and suggest that there are alternatives, to the degree that the strengths as well as the weaknesses are fully exposed, I think the power is going to remain. I think that is fundamental to the bill.

Educationally there is absolutely nothing in the bill that would cause the test to be poorer than they are now. From my perspective, everything in the bill would cause the tests to be far better than they have ever been if that is the direction that seems to come out of that public debate.

Mr. ANDREWS. May I ask you to comment? If I understand one of the arguments to the contrary, what good would it do? In other words, I understand the opponents to say that if this bill became law so that the applicants for the various schools, say, next year would not have the benefit of having the questions that are going to be submitted to them in advance of taking the test so that they might find the answers and hence if their memory is any good, make 100 or 1,000, whatever, in order to avoid that, I believe they say it would be necessary at each testing period to come up with a new test.

Do you concede that probably to be the case?

Dr. PERRONE. You have already heard from Mr. Rever that ACT has a new test right now. That does not seem to be an obstacle.

Mr. ANDREWS. What good would it do to make public the test given in 1979 so there would be this public debate and examination as to the credibility or feasibility or value of the 1979 test if the 1980 test is in fact going to be an altogether different test?

Dr. POTTINGER. It would not help those who took the 1979 test, but once we understand how the 1979 test was constructed, why the ones that are under dispute were put in, we may change the 1980 test.

Mr. ANDREWS. But you are not going to make the 1980 test. The same people who made the 1979 test are going to make the 1980 test.

Dr. POTTINGER. That is the point. They are going to make the decision every year without any of us having an input into the process.

Mr. ANDREWS. You cannot have an input into the process by having made available to you the 1979 test? They are going to make the 1980 test and you will not have a copy until it has already been given, right?

Dr. POTTINGER. That is right, you are correct. We would not have direct input into the item writing. They would still produce those tests. I think we would demonstrate whether the tests are in fact being effective or not.

Let's go back to a simple example. There may be items on the test that simply are not appropriate to what we are trying to test. We don't know that without seeing the items. That is, someone may think they are appropriate. The testers may feel they are appropriate and the people who wrote the items may feel they are appropriate. That is a very limited number of people who have control over the decisions that are being made.

We are simply asking that somehow the process be open so that more than these privileged few people make the decisions about how our young people are admitted or rejected from schools. That is the basic part of it.

Mr. ANDREWS. Mr. Erdahl?

Mr. **Erdahl.** I believe the spread and standard are in error, I think we had the figure pointed up of 100 points in the SAT test and information the staff has provided me, I understand it is a 30-point spread. Maybe that is too big, but it is not as big as 100 points.

Would you want to comment on that?

Dr. **Perrone.** It depends on where you are on the scale, if you are on the upper or lower end of the scale.

Mr. **Erdahl.** I think you made the point it was 100 points. Information we have, according to ETS, is that it is a 30-point spread.

Mr. **Weiss** [presiding]. I asked what is the range of error. He responded to what the range of error would be and I think that is correct.

Mr. **Erdahl.** Another point I would like to dispose of here, again with regard to the article in the Washington Post, since publication of any particular test would lead to its immediate retirement, the corrective examples would be helpful only in the same generalized way as the test study guide already available.

If we are trying to get at a problem of tests that probably don't reflect the applicant's true ability, I am not sure that the publication of the tests is the best way to do it. The test, I suppose at times, is a learning tool and hopefully our schools use them wisely in that way, but a test is also a diagnostic tool. It is not the treatment.

We have had a concern here today that maybe people are ill-equipped to survive in law school or medical school. Again, that is not the fault of the test. It is the fault of something in our educational system.

Or you might say that law school doesn't have the right requirements or the medical school has the wrong requirements. It seems like the testing doesn't get at that.

I am not sure what one achieves by requiring that the testing agency come up with new tests all the time.

Dr. **Pottinger.** Can I answer that with an analogy?

I think it is an old saw. If we say the automobile, going back to my manufacturing analogy, if we say the automobile manufacturer just makes the car and it is unsafe at any speed, maybe I should add that if the consumer buys the car or the dealer sells the car, that maybe we should put the onus of responsibility on the buyer and the dealer.

We are talking about a system, an organizational or social system, that operates. To say these people simply manufacture the tools that then others rely on appropriately or inappropriately gives them no sense of responsibility.

It is, I think, akin to saying, the same response we would get if we said any manufacturer of any product has no responsibility for it, it is only the user, the dealer or intermediary people.

Mr. **Erdahl.** I am not sure I follow your analogy. It seems to me I don't have to buy that car. I can take a bus or train or walk. The same is true, I think, with the tests. If somebody comes up with a better one, I think the real pressure should be on the educational institutions in the whole area of education.

If the tests are not good, let's come up with some better ones.

Dr. **Pottinger**. But we cannot come up with better tests until we understand. As long as the people who have control over the development of the tests that we claim without proper data are not good, we simply don't have the leverage to create innovations to come up with better tests. We cannot do that until we have open debate of what is now going on.

Mr. **Erdahl**. Then Mr. Rever has misinformed us that this information is available to everyone who ever took the test. If they have some memory of when they took the test and said, those are inappropriate tests or they are not relevant, it seems that information is available now.

Dr. **Pottinger**. It is there theoretically, not practically. Every student who has the gumption or finances to hire an attorney or take action, some appeal procedure or proper action, can probably get access to it. It is not a theoretical issue, but it is a practical issue of access to the information. We got access to the information after 2 years in the courts.

Mr. **Andrews** [presiding]. Have you seen the questions, all of them given on the LSAT for the last 15 years?

Dr. **Pottinger**. No, I have not.

Mr. **Andrews**. How many of them have you seen?

Dr. **Pottinger**. I have seen what has been published publicly, that is, what the ETS has released some years ago. They released a past test of the SAT so people could understand some of the items. A few of those items are in my testimony, the comparison-contrast-type items.

Mr. **Andrews**. But you have not seen the items on the SAT for 1978?

Dr. **Pottinger**. No, and ETS has the full authority to determine whether I can see those items or not and they have the full authority to decide whether I am a credible expert to view those items.

Mr. **Andrews**. You say you have not seen them?

Dr. **Pottinger**. I have not seen—I have seen past tests, the items that are public information.

Mr. **Andrews**. But you have not seen all the questions that were on the SAT or any of the SAT tests that were given in 1978?

Dr. **Pottinger**. That is right.

Mr. **Andrews**. Dr. Perrone, let me ask you this question: Do you subscribe to or support or believe in, adhere to the basic principle of testing?

Dr. **Perrone**. I think that testing has some possible uses.

Mr. **Andrews**. Do you think an institution of higher education should cause the applicants to be submitted to some written test?

Dr. **Perrone**. Not necessarily.

Mr. **Andrews**. Well, do you think they should have the prerogative of doing that or not doing that as the institution might choose?

Dr. **Perrone**. I think colleges and universities have probably not been as responsible as they might have been over the years in buying into this particular mode of testing.

Do I think colleges and universities ought to have some kind of an admissions process? Yes, I do.

Mr. **Andrews**. Do you think that should include requiring the applicant to submit to some written test?

Dr. **Perrone**. Again, I will say not necessarily. I think colleges and universities might make a range of decisions about that particular factor.

Mr. **Andrews**. Do you think the institution should have the prerogative to make that determination and require testing if they choose?

Dr. **Perrone**. I think they should have that opportunity and I hope they would take advantage of that opportunity within an atmosphere where there is full disclosure about those tests and what they are and what they mean and within the context that students have access to the decisions, the judgments that are being made about them, and that they will ask the question, have students been treated fairly in this process as well.

I will only say again, I think the fundamental issue in this bill goes all the way back. In some ways we have been all the way around it, we have taken societal issues, but there is a fundamental issue for students who take a test. Ought they not have the opportunity to review that test, the questions, the answers, and their responses. That, I think, is very fundamental in this bill. I think that is a right that a student ought to have in this process.

I think the whole testing enterprise would be far better for being open to that than to continuing an argument that somehow that will reduce validity or item quality or that some technical issue will be compromised. There is no evidence that that is true.

Mr. **Andrews**. I see.

Dr. **Rever**. It is hard to argue against Dr. Perrone's basic tenet, that is, that students ought to have the right to see their examination and its basic contents. That is the kind of appealing issue that I think everyone would love to endorse.

Our difficulty with that is that for one thing it is for no observable purpose whatever. Whatever learning will take place for the student, as you can see, as you pointed out in your line of questioning, is relatively small. The examinations really are a sample of behavior, learning. Which items one misses has no implication in general for whether or not they have mastered something very different from that, not even related to the question.

So, consequently, when you are talking about student rights, you are talking about protecting them from some kind of corporate power which is unwarranted. We suggest to you, in fact, Dr. Perrone's very own statement of a while ago argues against the passage of these bills on the basis of whether institutions over the past two decades have been studying and evaluating examination programs as they try to move to comply with the laws for equal opportunities in this Nation.

Bowdoin and Harvard and others have chosen not to employ the examination in their admissions programs. That is a very powerful argument that indeed our programs are responsive to the needs of those institutions and they are attending to their social responsibilities in evaluating the kinds of services we have.

So we look at the benefits. Are there really truly benefits? Indeed, researchers do look at bar examinations and try to improve them for us. We hire people for that very purpose to represent disciplines across the spectrum of American education and the secondary school systems. So really of what benefit is it to the student?

We submit there has been no clear need for that kind of right to be extended by our Government to the student at the moment.

Mr. ANDREWS. Do you have a concluding statement?

Mr. WEISS. Just a couple of questions, Mr. Chairman. Just a little bit ago Mr. Erdahl, in responding to Dr. Pottinger's automobile analogy, said that he should have the option of not using it, not buying the automobile at all, or using the bus or train or whatever else.

I think the problem is that the student may in fact have no choice. For example, let me put the question, the student who wants to apply to law school, what percentage of the law schools today will accept a student who does not take the LSAT?

Dr. PERRONE. I don't know any.

Dr. POTTINGER. I don't know the percentage.

Mr. WEISS. There are some, but probably miniscule as far as percentage is concerned, as far as the percentage of law schools that don't require the SAT. I don't know if any of you are familiar with the specifics.

In the first day of testimony that we have here, somebody referred to the fact that somehow within the last few years apparently there was an LSAT examination given and the professors of Georgetown Law School then examined the answers and questions and concluded that their answers in 25 percent of the instances would be different than that which the LSAT had in fact designated as being the correct answer.

Now it seems to me that if that kind of testing and testing floor were open to public scrutiny and public critique, that in all probability the LSAT people and the testing services generally would be much more concerned and careful about the kind of tests which they would be putting forward for the students of next year's LSAT's.

Isn't that really the concept behind what the benefit is to the student and to the society, Dr. Perrone?

Dr. PERRONE. I think that is the argument I have been posing throughout my testimony. I think Dr. Pottinger has posed the same issue, that the item quality will undoubtedly improve.

Mr. WEISS. In conclusion, I want to express my appreciation to all the witnesses on both sides of the issue for their participation. I think they have helped to further define, and your questions certainly have, what the issues are here.

In summary, I want to say that I do not nor does H.R. 4949 intend to eliminate standardized testing. What I want to do is make sure it is an open process, that people understand what is in it and that in fact we can have a better system that everybody can be happier with.

I thank you very much.

Mr. ANDREWS. I, too, on behalf of Chairman Perkins, would like to thank all of you for your very valuable contribution. Unless there is an objection, we will now adjourn to reconvene on September 24, 2 weeks from today, at which time we will consider these same bills.

Thank you.

[Whereupon, at 1:25 p.m. the subcommittee adjourned.]

TRUTH IN TESTING ACT OF 1979; THE EDUCATIONAL TESTING ACT OF 1979

MONDAY, SEPTEMBER 24, 1979

House of Representatives,
Subcommittee on Elementary, Secondary,
and Vocational Education,
Committee on Education and Labor,
Washington, D.C.

The subcommittee met, pursuant to notice, at 9:30 a.m., in room 2175, Rayburn House Office Building, Hon. Ted Weiss presiding.

Members present: Representatives Weiss, Kildee, Buchanan, and Erdahl.

Staff present: John F. Jennings, majority counsel; Jeffrey W. Brennan, staff intern; and Martin LaVor, minority senior legislative associate.

Mr. **Erdahl** [presiding]. Good morning. We will get our hearing underway.

I was asked to start the hearing this morning. Chairman Perkins is on his way in. I see Mr. Weiss is here. I will begin the hearing because we do have several witnesses to appear today.

This is the fourth day of hearings. Two more days are scheduled on H.R. 3564 and H.R. 4949.

We have an interesting and distinguished panel this morning. If they would take their places around the witness table this morning, we could get the hearing underway.

Even though some of our colleagues are not here yet, this will go into the record and we will establish the record this morning.

The people on the panel this morning are: Dr. Urbain J. DeWinter, director of admissions, College of Arts, Cornell University; Dr. Richard Berendzen, president, American University, representing the American Council on Education; Dr. James W. Loewen, director of research, Center for National Policy Review, Catholic University School of Law; Ms. Althea Simmons, director, Washington Office, NAACP; Dr. Lionel Newsom, member, board of directors, National Association for Equal Opportunity in Higher Education, president, Central State University; Mr. Sheldon R. Goldstine, Cleveland, Ohio; and Dr. I. Leon Smith, director, Professional Examination Service.

First I will call on Dr. DeWinter, and now I will relinquish the Chair to Representative Weiss.

Mr. **Weiss** [presiding]. Let me apologize. It used to be that only death and taxes were certain. I now have a third part of that. That is that the Eastern shuttle will be late going or coming from New York. I apologize for myself. That is the reason I am late.

Thank you very much.

Dr. DeWinter, if you would proceed.

STATEMENT OF URBAIN J. DeWINTER, DIRECTOR OF ADMISSIONS, COLLEGE OF ARTS, CORNELL UNIVERSITY

Dr. **DeWinter**. Good morning. My name is Urbain J. DeWinter, associate dean and director of arts and sciences admissions.

Let me give you personal background from which I approach the proposed legislation. I was a high school teacher for 3 years and have taught at the college level for the last 15 years. Since I obtained my Ph. D. in Spanish literature from the University of Pennsylvania in 1972——

Mr. **Weiss**. Dr. DeWinter, excuse me. That microphone is not as sensitive as it ought to be. Maybe you could pull it closer and down.

Dr. **DeWinter**. Thank you.

I have taught at Cornell University where I have served both as a faculty member and an administrator. I am currently associate dean and director of undergraduate admissions for the College of Arts and Sciences.

In this capacity it has been my responsibility since 1975 to direct the admissions program of the college and more specifically to direct the selection process for the approximately 8,000 students who apply for admission to this college.

I want to point out to the committee that I do refer specifically to the College of Arts and Sciences because at Cornell we have seven different undergraduate colleges (not to mention the several graduate divisions), each one varying somewhat in its policies and procedures.

My colleagues who serve as directors of admission in these colleges have for the most part academic backgrounds in such fields as the history of science, botany, architecture, or agricultural economics.

I mention these facts because there is no unanimous opinion on this question at Cornell, and the university as an institution has not taken an official stand on the proposed legislation.

I would also like to mention a little later an associate of ours, Prof. E. Wood Kelley from the Cornell Department of Government, will be available for additional testimony when he arrives later this morning, especially on the legislation in bill 3564. He is a member of the admissions committee.

As a member of the admissions committee, Professor Kelley was overall project consultant for the evaluation of the Elementary and Secondary Education Act for the House Education Committee prior to its recent reauthorization, and has consulted for the House Education Committee, the NIE, the National Opinion Research Center in testing, the Federal Judiciary Center of the Supreme Court, and other organizations.

I accepted the invitation to testify on the proposed bills because I believe that the passage of this legislation will reinforce our own selection policies and procedures which I think are fair and comprehensive, and it will ultimately increase our understanding of the tests, thereby placing them in their proper perspective, without destroying their validity.

I believe this legislation is in the interest of fairness to students, parents, teachers and in the overall interest of education.

There may be some inconveniences that arise as a result of this legislation, but these should be resolved in due time and are outweighed by the overall advantage to a fairer educational system.

In order to put the question of tests in the proper perspective, I want to say something about the philosophy of admissions as articulated by the faculty of the College of Arts and Sciences and how this philosophy is implemented.

It is our belief that faculty and students benefit both academically and personally from a diverse student body. The College of Arts and Sciences admits men and women who have diverse social, economic and cultural backgrounds, racial and national identities, and talents.

All students entering the college must be prepared to contribute to and benefit from the environment of Cornell. Therefore, admission selection is an evaluative process that considers each individual's achievements and potential for success at the university.

The most important criteria for admission are intellectual potential and commitment—a complex combination of ability, achievement, motivation, diligence and educational opportunity. Nonacademic qualifications are also important.

The college seeks individuals with outstanding personal qualities and a record of significant involvement in extracurricular endeavors. Students with unusual talents and achievements in music, acting, creative writing, science, athletics, politics, and other areas are given special consideration in the admission process.

A committee of faculty members and admissions officers reviews the materials submitted by the applicant. The folder includes the following items:

There are four different areas that are important.

One. The application form itself which contains a lot of information on the student's background, and his or her interests, accomplishments, and goals; for example, one of the questions is a 300-word essay that reads as follows: "Describe a problem that faces your home, school, city, region, or nation. How did the problem originate, and how might it be solved?" Or another question seeks to find out "What educational objectives, interests, experiences, career or professional goals or other factors have influenced your decision to apply for the program you have specified?"

Two. An individual school report which would demonstrate that the student has been prepared for a challenging college curriculum. This form contains the following statement: "A full and candid report from the school is essential if the applicant is to be given fair consideration. We therefore ask for careful ratings and comments about character and ability by a school official who knows the student well." The ratings and comments refer to class rank, transcripts, test scores, personal characteristics such as academic motivation and selfdiscipline, as well as overall performance of the candidate in extracurricular activities; also this form provides the opportunity to make additional comments pertaining to special circumstances that may be relevant to the student's application. This is important information because due consideration is given to the grading system, the courses available and taken, and cultural opportunities.

Three. Letters of reference are welcome when they provide an assessment of the applicant's academic or personal qualities as are other subjective reports such as the comments of an interviewer.

Four. Finally, we require the results of the SAT or ACT examinations and three achievement tests which are evaluated with the rest of the applicant's materials.

I would characterize our method of selection as a holistic one in which any one factor cannot easily be separated from the others. The complete application folder is always greater than the mere sum of its parts.

Furthermore, it is a method that looks positively for what the student has accomplished in the past and is likely to accomplish in the future rather than for what he or she has failed to do. It is an evaluation of each candidate in light of the opportunities and accomplishments he or she has demonstrated.

The role of standardized tests in this review has not been quantified. Standardized tests have been reviewed by our committees as one measure of reasoning abilities in both verbal and mathematical areas. As we understand them, the SAT's and achievement tests place the applicants from a wide variety of schools on the same scale in order to make some meaningful comparisons.

Clearly, we would not select our students on the basis of the tests because that would be in direct contradiction to our philosophy. To equate admission with the score on a test would be to equate all applicants with the characteristics of the test which would entirely eliminate our diversity.

Therefore, in our case, standardized tests have had a supplementary value rather than an absolute value, and this will continue to be the case. This does not mean that we reject the notion of a standardized test. For example, we use the regents examinations which are given in New York State.

But our committees have refrained from placing too much emphasis on the scores Instead they use multiple criteria in determining admission. Therefore, when speaking about our own method of selection I would say that whether the answers are disclosed or not, the current legislation is unlikely to have a strong impact on our procedures.

Nevertheless, I endorse the current proposal pertaining to disclosure of facts because there does exist the danger of abusing or misusing tests, as appears to be the case in some professional programs and institutions that receive a large volume of applications. Furthermore, they are frequently misunderstood or poorly appreciated by the general public.

One. Tests make it extremely easy to administer vast amounts of paper and material that can be digested in a highly mechanical fashion, with great speed and uniformity. We have to be careful with the very visible numbers lest they distort our vision and prevent us from taking the appropriate time to make a thorough review of each case. It is important to recognize the efficiency of clean-cut numbers for what it is worth.

Two. We cannot assume that what is generally true about society is also true about the individual and therefore good for him or her.

On the basis of statistical analysis, for example, we might say that the average-size family in the United States consists of 2.7 children. Does this imply that I should have 2 or 3 children, or 0.7 children?

We can be tempted to exclude individuals or groups of individuals from consideration for a program because they do not meet the average number. If that number becomes so significant that on that basis I am excluded from a program, then I as an individual have a right to see the test upon which I have been excluded.

Three. Objectivity is frequently misunderstood. We speak of objective tests or decisions, confusing one part with the whole. To make an objective decision does not mean limiting the factors I take into consideration to quantifiable areas. Subjective judgments are part of the objective decision. We can too easily confuse quality with an "objective" number, as in describing an individual or a group by means of a number.

Four. Disclosing the tests as proposed in H.R. 4949 would add to our knowledge of the test and demystify the secret contents that promote misunderstanding of the tests among students, parents, and admissions officials.

It would challenge all of us to become still more knowledgeable about the test, and to enable us to give a student or parent an explanation of just what a high or low test score represents, how it is interpreted by both the testing agency and the admissions office.

Disclosure would lead to a more realistic appraisal of the strengths as well as the limitations of a test. Ultimately it will strengthen the credibility of tests, for they will be examined by a wide variety of educators who are interested in maintaining their high quality. Undoubtedly intelligent people will express disagreement over both the questions and the answers of the tests.

The most important objections that have been raised to the disclosure of tests involve cost and equating. Ample testimony has been given in New York and Washington showing that the cost of test development represents but a relatively small fraction of the total cost of test administration.

Equating is often confused with validity. Equating simply means that we rank the students in some common order across test administrations. However, equating is a complicated psychometric procedure that should not be confused with the validity of the test which deals with the predictive value of the test. It is validity and not equating which is our main concern.

As I understand the bill, it was drafted in consideration of the testing industry's concern to exempt from disclosure questions used for equating and pretesting purposes.

In sum, I believe this legislation will provide us with more useful information about standardized tests, an opportunity to monitor the appropriate development, correction and administration of the tests, and continue to encourage the responsible use of tests in schools and admissions office.

Mr. WEISS. Thank you very much. We will hold off the questioning until the entire panel has had occasion to make their statements and then we will proceed to questions.

Now we will proceed in the order in which the witnesses appear on the list. Our next witness is Dr. Richard Berendzen.

STATEMENT OF RICHARD BERENDZEN, PRESIDENT, AMERICAN UNIVERSITY, REPRESENTING THE AMERICAN COUNCIL ON EDUCATION, ACCOMPANIED BY SHELDON STEINBACH AND DOUGLAS WHITNEY OF THE AMERICAN COUNCIL ON EDUCATION

Dr. **Berendzen.** Mr. Chairman and members of the subcommittee, my name is Richard Berendzen. I am university provost and president-elect of the American University here in Washington, D.C.

I thank you for this opportunity to offer testimony in behalf of the American Council on Education and seven other major education associations on H.R. 3564 and H.R. 4949.

With me today are Mr. Sheldon Steinbach and Mr. Douglas Whitney, of the American Council on Education.

As you may know, the membership of the eight organizations we represent here today includes nearly all the colleges and universities in the Nation. And the American Council on Education serves as the primary coordinating agency for the Nation's higher education community.

Thus, our testimony today represents generally the concerns and interests of the diverse postsecondary institutions that use standardized tests in their admissions processes, as well as the present and future students of these educational institutions.

We recognize that faculty, students, parents, and others have a genuine and legitimate interest in the nature and use of the standardized tests that play a role in the admissions process at many higher education institutions.

During the past four decades, such tests have proven to be useful at many institutions in assessing students' potential for completing undergraduate, graduate, and professional programs of study.

In many instances, test results have been especially helpful to students who attended schools and colleges not well known to admissions committees and to those whose prior educational work did not reflect their full potential for further study.

At the outset we wish to stress that while acknowledging the interests of students and others in the nature and use of such tests and concurring with some of the stated goals of the proposers, we believe that Congress should not enact such legislation at this time.

Portions of each proposal now before the House are unnecessary in that they simply codify the current practices of most testing programs and make no significant changes or improvements.

Further, the proposers of H.R. 4949—which would require the disclosure of test questions and answers after each test administration—have not proven the need for this major change in test program procedures on the basis of demonstrated educational benefits to students, on a definitive analysis of relative costs and benefits to students or institutions, or on well-documented instances of improper or harmful past practices.

Finally, the portions of H.R. 4949 that require test agencies to make public all studies conducted on behalf of individual colleges and the portion of H.R. 3564 that prohibits the consideration of the test results of other applicants in determining whether or not an individual should be admitted to an institution would severely limit the ability of colleges to develop admissions procedures that best serve the needs of both students and institutions.

Historically, Congress has shown an admirable reluctance to legislate matters at the heart of education and curricula. College admissions procedures—including the evaluation of test results—are an outgrowth of and a reflection of the mission and goals of each institution. The information required from each student applicant properly reflects the curriculum and requirements of the institution and educational program to which application for admission is made.

The subject of the proposed legislation is an educational matter and properly the province of the faculty of each institution. Although the proposers disclaim any intent to interfere with admissions procedures, it is our view that the evolution of subsequent regulatory and legislative actions will do precisely that by moving from one factor in the admissions process—standardized test information—to others.

I now turn to a brief analysis of the proposed legislation.

First of all, legislation in areas of information and privacy is unnecessary. The enactment of legislation in any area not previously subject to legislation suggests the existence of an undesirable and intolerable condition that can and must be remedied by law.

Do such conditions exist in the area of educational admissions testing? We think not and do not believe the proponents of the bills now before you have carried their burden of demonstrating a clear need for such legislation.

Today, qualified students have greater access to higher education in the United States than ever before. More than 90 percent of the students currently seeking admission to undergraduate study are admitted to the college of their first choice. More than 95 percent are admitted to some college.

As the number of 18-year-olds declines, more and more undergraduate institutions can be expected to use test results for inclusive purposes to help identify and admit all students able to complete undergraduate programs of study rather than to exclude those who appear less likely to do so.

At the graduate and professional levels, most institutions have already recognized the desirability of broadening their student bodies and have achieved diversity through vigorous programs of affirmative action, special academic assistance and multiple criteria for admissions. This has occurred in the absence of legislation concerning the use of admissions tests.

Both H.R. 3564 and H.R. 4949 prescribe specific items of information about each admissions test that must be provided to students when they register for the examination. While we support the practice of providing students with useful information, it is our judgment that nearly all the required information is already provided to students.

The United States Student Association conceded as much in their testimony concerning H.R. 3564. The fact that section 3 of H.R. 3564 and section 6 of H.R. 4949 essentially require testing programs to continue to provide students with the information now available is hardly an argument that demonstrates a need for the legislation.

Finally, no persuasive evidence has been presented to suggest that test agencies are now handling student test records improperly. To the best of our knowledge, test results are currently sent in identified form only to institutions and agencies designated by the student—either by a specific request or by a general permission applying to scholarship programs, special information services and the like.

Some persons have been critical of the College Board and Educational Testing Service for refusing to release identified test results to the Federal Trade Commission for use in its study of coaching schools. Section 6 of H.R. 4949 would require them to refuse other similar requests.

Second, requiring the disclosure of test questions is unwise. Section 6 of H.R. 4949 requires the release of test questions and answers to students and to the public after their first use. This would effect a major change in the current operations of all educational admissions testing programs.

This requirement is the major point of contention surrounding H.R. 4949. That this would result in an abrupt change for testing programs is not a sufficient reason for us to oppose disclosure. It is the anticipated effects and highly questionable benefits of this seemingly simple change that cause us to oppose this section.

It is generally acknowledged that the disclosure provision would necessitate the development of more test questions and more test forms than are currently needed. It would be impossible to reuse questions that have been shown to be useful and valid.

It would be unwise to legislate a change of this magnitude without adequate assurance that the quality of information gained from admissions tests will not thereby be compromised without a clear indication of the number and identity of the students who might benefit from such changes and without a comprehensive estimate of the associated costs.

It is by no means clear that there will be any significant educational benefit for students from the required disclosure of test questions and answers. Experience suggests that the use of a pretest or the development of more extensive diagnostic information to accompany student score reports or both would better accomplish the educational purposes claimed for H.R. 4949.

We note that the learning that occurs for students whose questions and answers are returned is a matter that could be rather simply studied. This should be done before disclosure is required in order to help us all understand more fully the implications of the legislation.

The possibility that the requirements in section 5 of H.R. 4949 would result in a decline in the quality of information that is gained from these tests is a major concern. Although some persons have testified that quality will not be adversely affected, we believe that most experienced test experts would disagree or express doubts that this will be the case.

The possibility that test information will be less useful than it is now is particularly vexing in light of the failure to demonstrate that students will benefit from the passage of this legislation.

The number of students who would request copies of their test questions and answers is not known. Without a satisfactory estimate of the number of students who would do so, it is impossible to evaluate the societal benefits that would result from enacting H.R. 4949.

If few requests are made, the marginal costs of providing copies will prove to be high and the nominal charges permitted will be beyond the ability of low-income students to pay.

You should be aware that many institutions ask students to complete standardized tests chiefly to assist in providing guidance in matters of course selection and planning. Some test agencies—for example, the American college testing program—currently make retired test forms available to institutions for use in student orientation or preregistration testing programs on the campuses.

It seems likely that if the disclosure requirement is enacted, secure test forms will no longer be available for this purpose or that they will be much more expensive to use for these purposes.

It should be noted that educational admissions tests are not all alike. It seems inevitable that the disclosure requirement would be more difficult to effect for some tests than for others. No evidence has been presented that suggests that any admissions testing program can be operated in this manner, let alone all of them.

This legislation is untimely. During the past year, the role of ability tests in American education and society has been under study by a broadly representative panel of educators and others. This study by the National Academy of Sciences' Committee on Ability Testing is scheduled to be released in midyear 1980.

Although we have no knowledge of the conclusions or recommendations to be included in the report, it seems likely that considerable attention will be given to the use of tests in educational admissions. Action on testing legislation before this report is released would prejudge the evidence and preempt the conclusions of the study panel.

No crisis exists to warrant hasty action on these proposals before the committee's report is released. We expect the report will contribute much to the discussion of issues associated with this legislation.

And, finally, each year more than 1,000 studies of test validity and related matters are conducted by or on behalf of individual colleges and universities. Each involves students enrolled at a single institution and each examines the usefulness of admission test results in conjunction with other relevant information about the students as they relate to indicators of program completion and achievement at that institution.

Currently, the results of these studies are available to faculty and admissions officers at that institution for their use in developing appropriate admissions procedures. None is of interest beyond that institution. None necessarily has relevance for students at other institutions.

Yet, section 4 of H.R. 4949 requires that all be made public if they have been conducted by a test agency on behalf of the institution or if they have used data obtained from a test agency.

Individual institutions should be encouraged to conduct such studies. The public filing requirement of section 4 of H.R. 4949, however, is likely to discourage these efforts. The test agency studies prepared for an institution consist largely of tabular material.

They lack the interpretive context that is known to admissions committees, but that would be unknown and quite possibly misunderstood by other readers. It would serve no purpose to require the publication of studies that pertain to single institutions.

I now wish to comment on the matter of improvements in educational testing: What have been and are alternatives to Federal legislation with its attendant risks?

As a general principle, we believe that legislation should be used as a remedy only when less drastic alternatives have been exhausted. Historically, three nonlegislative means have proven useful in identifying problems and effecting needed improvements in educational testing:

One. Evaluations of tests, voluntarily agreed to by the test sponsors, conducted by qualified persons with no vested interest in the outcome,

Two. Standards for good practice that clarify the procedures for developing and using tests and for preparing test manuals that are generally endorsed by professional test experts, and

Three. Publication of analyses of issues and recommendations prepared by broadly representative ad hoc committees of educators and others.

I call your attention to the examples of these efforts that we have detailed in the written testimony we have submitted.

These examples of significant efforts to guide and improve the use of educational tests stand as evidence of the education community's intent to keep its house in order. These activities have led to changes and improvements in educational testing. They are read and understood by persons involved in institutional admissions decisions.

In the course of these hearings, witnesses have raised a number of serious allegations about deficiencies in current admissions testing programs. We believe that all can and should be addressed in one of the above ways and that these nonlegislative approaches will prove more effective at identifying and remedying deficiencies than would any legislative solution.

I believe I can say in all candor that the postsecondary education community welcomes the public debate and interest that State legislation and the two bills now before the Congress have prompted.

We sincerely believe that both test producers and test users have policed their own practices well and have devoted enormous resources and effort to develop the art of educational testing. And, we remain committed to continuing that course in the future. In so doing, we welcome the active interest and assistance of the concerned groups who represent the interests of test takers.

Specifically, the American Council on Education makes commitments to future progress as outlined in the formal statement.

Mr. WEISS. Thank you very much, Dr. Berendzen.

[The complete statement of Dr. Berendzen follows:]

PREPARED STATEMENT OF DR. RICHARD BERENDZEN, UNIVERSITY PROVOST AND PRESIDENT-ELECT, THE AMERICAN UNIVERSITY, ON BEHALF OF THE AMERICAN ASSOCIATION OF STATE COLLEGES AND UNIVERSITIES; AMERICAN COUNCIL ON EDUCATION; ASSOCIATION OF AMERICAN UNIVERSITIES; ASSOCIATION OF CATHOLIC COLLEGES AND UNIVERSITIES; COUNCIL OF GRADUATE SCHOOLS OF THE UNITED STATES; NATIONAL ASSOCIATION FOR EQUAL OPPORTUNITY IN HIGHER EDUCATION; NATIONAL ASSOCIATION OF INDEPENDENT COLLEGES AND UNIVERSITIES; AND NATIONAL ASSOCIATION OF STATE UNIVERSITIES AND LAND-GRANT COLLEGES

Mr. Chairman and Members of the Subcommittee, thank you for the opportunity to testify, on behalf of the 8 Higher Education Associations, concerning admissions testing in higher education and to discuss the proposals for legislation in this area (H.R. 3564 and H.R. 4949). As organizations whose members include nearly all of the colleges and universities in the nation, we are troubled by these proposals and concerned about the impact their enactment would have on the admissions process at colleges and universities.

SUMMARY STATEMENT

We recognize that faculty, students, parents and others have a genuine and legitimate interest in the nature and use of the standardized tests that play a role in the admissions process at many higher education institutions. During the past four decades, such tests have proven to be useful at many institutions in assessing students' potential for completing undergraduate, graduate and professional programs of study. In many instances, test results have been especially helpful to students who attended schools and colleges not well known to admissions committees and to those whose prior educational work did not reflect their full potential for further study.

At the outset, we wish to make it clear that, while acknowledging the interests of students and others in the nature and use of such tests and concurring with some of the stated goals of the proposers, we believe that Congress should not enact any legislation at this time. Portions of each proposal now before the House are unnecessary in that they simply codify the current practices of most testing programs and make no significant changes or improvements. Further, the proposers of H.R. 4949—which would require the disclosure of test questions and answers after each test administration—have not demonstrated the need for this major change in test program procedures on the basis of demonstrated educational benefits to students, on a definitive analysis of relative costs and benefits to students or institutions or on well-documented instances of improper or harmful past practices. Finally, the portions of H.R. 4949 that require test agencies to make public all studies conducted on behalf of individual colleges and the portion of H.R. 3564 that prohibits the consideration of the test results of other applicants in determining whether or not an individual should be admitted to an institution would severely limit the ability of colleges to develop admissions procedures that best serve the needs of both students and institutions.

Historically, Congress has shown an admirable reluctance to legislate matters at the heart of education and curriculum. College admissions procedures (including the evaluation of test results) are an outgrowth of and a reflection of the mission and goals of each institution; the information required from each student applicant properly reflects the curriculum and requirements of the institution and educational program to which application for admission is made. The subject of the proposed legislation is an educational matter and properly the province of the faculty of each institution. Although the proposers disclaim any intent to interfere with admissions procedures, it is our view that the evolution of subsequent regulatory and legislative actions will do precisely that by moving from one factor in the admissions process (standardized test information) to others.

ANALYSIS OF THE PROPOSED LEGISLATION

Legislation in areas of information and privacy is unnecessary

The enactment of legislation in any area not previously subject to legislation suggests the existence of an undesirable and intolerable condition that can and must be remedied by law. Do such conditions exist in the area of educational admissions testing? We think not and do not believe the proponents of the bills now before you have carried their burden of demonstrating a clear need for such legislation.

Today, qualified students have greater access to higher education in the United States than ever before. More than 90 percent of the students currently seeking admission to undergraduate study are admitted to the college of their first choice;

more than 95 percent are admitted to some college. As the number of 18-year olds declines, more and more undergraduate institutions can be expected to use test results for inclusive purposes to help identify and admit all students able to complete undergraduate programs of study rather than to exclude those who appear less likely to do so. At the graduate and professional levels, most institutions have already recognized the desirability of broadening their student bodies and have achieved diversity through vigorous programs of affirmative action, special academic assistance and multiple criteria for admissions. This has occurred in the absence of legislation concerning the use of admissions tests.

Both H.R. 3564 and H.R. 4949 prescribe specific items of information about each admissions test that must be provided to students when they register for the examination. While we support the practice of providing students with useful information, it is our judgment that nearly all the required information is already provided to students. The United States Student Association conceded as much in their testimony concerning H.R. 3564. The fact that section 3 of H.R. 3564 and section 6 of H.R. 4949 essentially require testing programs to continue to provide students with the information now available is hardly an argument that demonstrates a need for the legislation.

Finally, no persuasive evidence has been presented to suggest that test agencies are now handling student test records improperly. To the best of our knowledge, test results are currently sent in identified form only to institutions and agencies designated by the student—either by a specific request or by a general permission applying to scholarship programs, special information services and the like. Some persons have been critical of the College Board and Educational Testing Service for refusing to release identified test results to the Federal Trade Commission for use in its study of coaching schools; section 6 of H.R. 4949 would require them to refuse other, similar requests.

Requiring the disclosure of test questions is unwise

Section 6 of H.R. 4949 requires the release of test questions and answers to students and to the public after their first use. This would effect a major change in the current operations of all educational admissions testing programs. This requirement is the major point of contention surrounding H.R. 4949. That this would result in an abrupt change for testing programs is not a sufficient reason for us to oppose disclosure; it is the anticipated effects and highly questionable benefits of this seemingly simple change that cause us to oppose this section.

It is generally acknowledged that the disclosure provisions would necessitate the development of more test questions and more test forms than are currently needed; it would be impossible to reuse questions that have been shown to be useful and valid. It would be unwise to legislate a change of this magnitude without adequate assurance that the quality of information gained from admissions tests will not thereby be compromised, without a clear indication of the number and identity of the students who might benefit from such changes, and without a comprehensive estimate of the associated costs.

It is by no means clear that there will be any significant educational benefit for students from the required disclosure of test questions and answers. Experience suggests that the use of a "pre-test" or the development of more extensive diagnostic information to accompany student score reports, or both, would better accomplish the educational purposes claimed for H.R. 4949. We note that the learning that occurs for students whose questions and answers are returned is a matter that could be rather simply studied. This should be done, before disclosure is required, in order to help us all understand more fully the implications of the legislation.

The possibility that the requirements in section 5 of H.R. 4949 would result in a decline in the quality of information that is gained from these tests is a major concern. Although some persons have testified that quality will not be adversely affected, we believe that most experienced test experts would disagree or express doubts that this will be the case. The possibility that test information will be less useful that it is now is particularly vexing in light of the failure to demonstrate that students will benefit from the passage of this legislation.

The number of students who would request copies of their test questions and answers is not known. Without a satisfactory estimate of the number of students who would do so, it is impossible to evaluate the societal benefits that would result from enacting H.R. 4949. If few requests are made, the marginal costs of providing copies will prove to be high and the "nominal" charges permitted will be beyond the ability of low-income students to pay.

You should be aware that many institutions ask students to complete standardized tests chiefly to assist in providing guidance in matters of course selection and planning. Some test agencies (e.g., the American College Testing Program) currently make retired test forms available to institutions for use in student orientation or pre-registration testing programs on the campuses. It seems likely that secure test forms will no longer be available for this purpose or that they will be much more expensive to use for these purposes if the disclosure requirement is enacted.

It should be noted that educational admissions tests are not all alike; it seems inevitable that the disclosure requirement would be more difficult to effect for some tests than for others. No evidence has been presented that suggests that any admissions testing program can be operated in this manner—let alone all of them.

This legislation is untimely

During the past year, the role of ability tests in American education and society has been under study by a broadly representative panel of educators and others. This study by the National Academy of Sciences' Committee on Ability Testing is scheduled to be released in midyear 1980. Although we have no knowledge of the conclusions or recommendations to be included in the report, it seems likely that considerable attention will be given to the use of tests in educational admissions. Action on testing legislation before this report is released would prejudge the evidence and preempt the conclusions of the study panel. No crisis exists to warrant hasty action on these proposals before the Committee's report is released; we expect the report will contribute much to the discussion of issues associated with this legislation.

It is unwise to attempt to regulate the publication of research results

Results from studies of test validity that are of general interest and applicability should be published and widely distributed. Section 4 of H.R. 4949 seems an ineffective and objectionable way to accomplish this end. Frankly, we find the attempt to require that researchers associated with admissions testing programs publish all of their work at odds with the protections of academic freedom that are universally accorded their colleagues associated with universities. Certainly, distinctions must be made between preliminary and final reports; between memoranda and more extensive treatments or analyses; between papers prepared to stimulate discussion and those that present conclusions resulting from the discussions. This is not to suggest that an altered version of section 4 would be acceptable; it simply suggests that legislative solutions in the area of research publications are unenforceable as well as unwise.

Each year more than 1000 studies of test validity and related matters are conducted by or on behalf of individual colleges and universities; each involves students enrolled at a single institution and each examines the usefulness of admissions test results in conjunction with other relevant information about the students as they relate to indicators of program completion and achievement at that institution. Currently, the results of these studies are available to faculty and admissions officers at that institution for their use in developing appropriate admissions procedures. None is of interest beyond that institution; none necessarily has relevance for students at other institutions. Yet, section 4 of H.R. 4949 requires that all be made public if they have been conducted by a test agency on behalf of the institution or if they have used data obtained from a test agency.

Individual institutions should be encouraged to conduct such studies. The public filing requirement of section 4 of H.R. 4949, however, is likely to discourage these efforts. The test agency studies prepared for an institution consist largely of tabular material. They lack the interpretive context that is known to admissions committees but that would be unknown, and quite possibly misunderstood, by other readers. It would serve no purpose to require the publication of studies that pertain to single institutions.

IMPROVEMENTS IN EDUCATIONAL TESTING

As a general principle, we believe that legislation should be used as a remedy only when less drastic alternatives have been exhausted. Historically, three non-legislative means have proven useful in identifying problems and effecting needed improvements in educational testing:

(1) Evaluations of tests, voluntarily agreed to by the test sponsors, conducted by qualified persons with no vested interest in the outcome,

(2) standards for good practice that clarify the procedures for developing and using tests and for preparing test manuals that are generally endorsed by professional test experts, and

(3) publication of analyses of issues and recommendations prepared by broadly representative ad hoc committees of educators and others.

Examples of the first include the "Mental Measurements Yearbooks," published for 40 years by Professor O. K. Buros. The Yearbooks include critical reviews and complete research bibliographies for nearly all standardized tests. In addition, brief test reviews appear regularly in the professional journals of many education associations. Examples of the second include the Standards for Educational and Psychological Tests (1974) prepared by the American Psychological Association, the American Educational Research Association, and the National Council on Measurement in Education—national associations of individual researchers and test experts. Another is the statement on Responsibilities of Users of Standardized Tests (1978) published by the American Personnel and Guidance Association—a national association of individual counselors and guidance personnel. Each testing program issues explanatory and advisory materials that describe the tests and identify proper and improper uses. Examples of the third include the extensive report on the recent decline in scores on the scholastic aptitude test sponsored by the College Board, the comprehensive analysis of testing in education and society sponsored by the Russell Sage Foundation, and the upcoming report by the National Academy of Sciences Committee on Ability Testing.

These examples of significant efforts to guide and improve the use of educational tests stand as evidence of the education community's intent to keep its house in order. These activities have led to changes and improvements in educational testing; they are read and understood by persons involved in institutional admissions decisions.

In the course of these hearings, witnesses have raised a number of serious allegations about deficiencies in current admissions testing programs. We believe that all can and should be addressed in one of the above ways and that these non-legislative approaches will prove more effective at identifying and remedying deficiencies then would any legislative solution.

FURTHER COMMITMENTS TO PROGRESS

Portions of both H.R. 3564 and H.R. 4949 concern the nature of the information that is available to students when they register to take admissions tests and when they receive the report of their results. It has been said, during the hearings, that the current efforts of testing programs in this regard are insufficient and ineffective. We believe that a thorough and comprehensive analysis of current publications and practices in this area would identify any serious shortcomings in the information provided to students and that the suggestions and recommendations that would result from such a study would be influential with test sponsors. To that end,

> We invite organizations that represent the interests of students, parents, admissions officials, counselors, psychologists and test experts to join us in conducting and publishing the results of a review of the current student information booklets and test score reports prepared by educational admissions testing programs.

An effective program for self-regulation requires both a clear statement of intended goals and a continuing mechanism by which activities can be coordinated, maintained, and evaluated. To these ends,

> We will develop a clear statement of acceptable and desirable practices related to the development and use of educational admissions tests. We invite other interested organizations to join us in this effort. We believe that such a statement would result in any needed improvements in both areas, and we will convene a "Blue Ribbon" Committee representing the concerns of students, teachers, parents, colleges and universities, test experts, test sponsors, and the public. We will charge the committee to develop effective and continuing means for insuring sound practices in the development and use of educational admissions tests. This committee will meet for the first time shortly after the committee on ability testing report becomes available for study.

A SUGGESTION TO TEST SPONSORS

In developing test questions, the threats of ambiguity, bias, and faulty reasoning are always present—whether the writer is a teacher, scholar, or professional item writer. All test agencies conduct a review of admissions test items prior to their first use with students in order to minimize the possibility that these threats to test validity will obscure the actual abilities of the students. The review conducted by persons familiar with the test content and skills is an important step in the development of quality examinations, but many persons are not familiar with the procedures used or with the qualifications of the reviewers. Some of the feelings, expressed at these hearings, that admissions testing programs are shrouded in mystery and secrecy would be changed, we believe, if test sponsors prepared and distributed detailed descriptions of the review process used in developing their tests.

In addition, the possibility of involving qualified reviewers nominated by national associations of teachers and faculty should be explored.

Mr. WEISS. Next, Dr. Loewen.

STATEMENT OF JAMES W. LOEWEN, DIRECTOR, RESEARCH CENTER FOR NATIONAL POLICY REVIEW, SCHOOL OF LAW, CATHOLIC UNIVERSITY OF AMERICA

Dr. LOEWEN. Thank you.

I am the director of research at the Center for National Policy Review, Catholic University Law School. I have been in that position for 9 years and its primary purpose is to critique the Federal Government to make sure that the opportunities are maximized for minorities and for disadvantaged Americans.

Prior to my coming to the center, I spent 3 years as associate professor at the University of Vermont. I am still on that faculty on leave. I spent 7 years as a faculty member at Tougaloo College, Miss., a predominantly black institution.

I have also taught at Millsaps and at Harvard University, where I obtained my doctorate.

I say that to you because my testimony, which is not going to repeat my statements which you have, is based on my personal experiences with tests as well as my students' experiences, as well as my own fairly extensive research in the area. That experience is relevant because I have taught rural students, white and black, and urban. Based on that experience I want to highlight three points for you this morning.

I think there is a bias built into the aptitude tests on which admissions so largely depend. I think that bias is rural and urban. I think it is regional. I think it is racial and it is based on social class.

My second point—and I will return to each of these points in a minute—is that bias is involved in the tests, not just ineptitude distributed by region or by race or by social class.

I am going to give you an example of that bias.

My third point is that it is very difficult for me as a social scientist, as a sociologist in the field of educational sociology, to get data to do research in this area.

As you can guess, I am testifying on behalf of H.R. 4949.

Not only is it very difficult to get items from the tests, it is very difficult to obtain information from the college board of the ETS data as to how blacks perform, how whites perform, how various people in various States perform en masse.

Let me come back to the three points. The first point is bias. I assert there is a vast bias in the test. If you have my additional statement maybe you could turn to pages 2 and 3. This is a table which shows the cutoff point on the preliminary scholastic aptitude test, which is an ETS exam, the cutoff point used by the National Merit Scholarship Corp. which operates, as you probably know, a massive national merit scholarship program.

The merit program wants its competition to be national, and in order for it to be national it has to institute a different cutoff point for every State because, otherwise, its competition would be dominated by people from a few schools in Connecticut, Princeton, New Jersey, and a few other places in the United States.

So for the last two decades, at least, it has had a different cutoff point for each State and typically Mississippi has the lowest cutoff, and Connecticut the highest.

That is very interesting to me as a sociologist because, of course, I know Connecticut is the richest State in the United States in terms of per capita income, and Mississippi the poorest.

So I did a simple correlation of per capita income and PSAT cutoff point as used for the National Merit Scholarship Corporation. When you look at it you see that Alabama ranks 49th in income in the 50 States plus the District of Columbia, and it is 47th in the PSAT cutoff. Alaska is fourth in income and ninth, or tied for ninth in PSAT cutoff. That is a very close correlation.

The fact that Alaska has a little lower performance on the PSAT might relate to its ruralism or the distance from Princeton where this test is made up.

Kentucky, which I do not pick at random, is 43rd in income and 45th in PSAT cutoff—almost identical. When you do what is called the Spearman rank order correlation coefficient on the data as a whole, you come up with the astonishing coefficient of 0.83. They can rank in magnitude from zero to one. If they are zero that means there is no relationship between income and PSAT score. If they are one, that means they are a perfect relationship.

If you know income, you can perfectly predict PSAT cutoff score.

A correlation of 0.83 is one of the highest correlations you are likely to find.

There is a tremendous relationship between income and aptitude. There is a tremendous relationship between region of the United States and aptitude. When I first went to Mississippi, I found out that in both black and white schools a score of 400 or GRE was considered good and qualified you for college.

At Harvard most students had scores above 700. Am I to infer Mississippi is inept? I submit that is not the case. There is bias built into the tests.

My Tougaloo black students were drastically affected by this bias. The previous witness to me stated that 90 percent of all students get into the college of their first choice. I submit to you that is a meaningless statement. There is a chilling effect on your choice that comes from your score. Many students get a score on the LSAT or later when they are going into graduate school on the GRE, and they know not to apply to the college of their first choice. There is a chilling effect going on the score at all.

I have known black students who got terrible scores on the LSAT and, therefore, gave up. They are not getting a career of their first choice, let alone a college of their first choice.

First of all, to my original statement there is an appendix "A Clearinghouse of Civil Rights Research," which is a publication I edit at the Center for National Policy Review. (Retained in subcommittee files.) It contains two articles, one on IQ testing and one called "The Vicious Circle," by myself, which makes this argument clear, I think.

To summarize, Aptitude tests are misnamed. Aptitude correlates with socioeconomic advantage. Therefore, aptitude testing completes a vicious circle. Socioeconomic advantage leads to socioeconomic advantage.

Let me also ask you to turn the page to page 5 of my additional statement. This is titled "The Loewen Low Aptitude Test." This is a device I made up to show people the bias built into aptitude and IQ tests as given in this country. Number three is an actual item as remembered and modified by me from the test required by many graduate schools. I took it in the Midwest and I had no idea what I was doing on this item. I didn't know what was going on. I finally realized sake is a pronoun when you say for the sake of, and a pronoun cannot be an analogy.

Later, having experienced sake, and also China tea, but I had no idea that word was sake. Once you realize the word is sake then you realize the reasoning involved in an aptitude test. A Japanese sword play is analogous to sake as opera is to chianti. That is trivial.

But it depends upon an urban upper class vocabulary which I didn't have and which my students haven't had.

Item two directly above it is my own item which I would be happy to talk about. It is based on carpentry terms. My question is why is the sake item included on the test but not the carpentry item? I think the answer is clear: because the testmakers know about sake and don't know about carpentry; therefore, the student who knows about carpentry but doesn't know about sake is disadvantaged on the test.

I have some other words: argentum, Zeus, baroness, misanthrope, cuneiform, Hyden, Runnymede, Michelangelo, and on and on.

There is a serious bias built into the tests.

My final point is it is very difficult to get data. Not only is it hard to get items so that the student can't learn what he missed; it is also hard to get items that we can check easily for bias. We can't do this research on varied populations and the effect of aptitude tests on them. It is also difficult, and I see no reason for this, to get information on the breakdown scores as groups by race, by State and by social class.

ETS and the College Entrance Board have told me that they have a policy against providing racial data so they operate in secret. Although the college letterhead says it operates to expand educational opportunities, their tests deny equal educational opportunities to blacks and other minorities, to working class Americans, to rural persons and to some southerners and westerners.

In my additional statement I submit to you some specific suggestions for improving H.R. 4949 but in general I feel that bill would play a key role in promoting openness and thus inducing bias in the tests and use of tests.

I quote from Educational Testing Service in its description of the GRE:

> The aptitude test, like other standardized tests, makes it possible to compare students with each other regardless of their individual backgrounds. A GRE score of 500, for example, has the same meaning whether earned by a student from a small private liberal arts college or by a student at a larger public university.

This is a statement they have never submitted. It encourages the misuse of tests by an admissions office. Unless enlightened, as at Cornell, he is likely to say: The student got only 385; we cannot admit her. That happened to a student of mine.

My hope is that through the application of H.R. 4949 the American people as a whole will come to realize that statement is false and that aptitude tests have to be taken in the context of the situation in which the test is taken and the background of other applicants from that area.

Mr. WEISS. Thank you, Dr. Loewen. Your prepared statement will be, without objection, entered in full in the record and that will be the case for all the witnesses.

You may feel free to summarize if you so desire.

[The prepared statements of Dr. James W. Loewen follow:]

PREPARED STATEMENT OF DR. JAMES W. LOEWEN, DIRECTOR OF RESEARCH, CENTER FOR NATIONAL POLICY REVIEW, SCHOOL OF LAW, CATHOLIC UNIVERSITY OF AMERICA, AUGUST 15, 1979

"Truth in Testing" and other issues related to the use of standardized testing in education and employment have long concerned me, both as a researcher in sociology and as a college teacher. My research and teaching experiences in black and white America, rural and urban, have led me to several conclusions which I wish to share with the Subcommittee and its staff as you ponder these important matters. In this statement I will first summarize my experience and qualifications in this area, then describe some insights I gleaned from my own personal encounters with tests and from those of my students, and finally describe some problems in the research literature and in doing scientific research in this area.[1]

Experience and qualifications [2]

I hold a Ph.D. from Harvard in sociology with specialization in race relations, sociology of education, and field methods of research. These three areas are directly relevant to the topic of testing and to the research that has been done in this area, particularly because field methods have been the Achilles heel of the testing movement. (More on that later.) I taught for seven years at Tougaloo College in Mississippi, an excellent predominantly black college, and am now on leave from the faculty of the University of Vermont. At those institutions and in teaching positions at Millsaps College in Mississippi and Harvard University I encountered diverse sets of students, including rural Mississippians and rural Vermonters, Bostonians and students from New York and Philadelphia suburbs, children of millionaires and of sharecroppers, and of course blacks and whites.

Work in the sociology of race relations resulted in my sharing the first annual Spivack Award of the American Sociological Association in 1978. My recent work specifically related to testing includes three articles: "Stratification: Class Biases in IQ Tests," in Eighty-one Techniques for Teaching Sociological Concepts (Washington: American Sociological Association, 1979), "Breaking the Vicious Circle," in the Clearinghouse for Civil Rights Research (Vol. VI No. 1-2, 1978), and "Introductory Sociology for the Privileged: Four Classroom Exercises," in Teaching Sociology (Vol. 6 No. 3, 1979). As Director of Research at the Center for National Policy Review, I have planned a research project with Dr. William Sedlacek, Director of Research at the Counselling Center of the University of Maryland, titled "Possible Causes of Lower Black Scores on Aptitude Tests." I have also been campuswide Colloquium Speaker on "Aptitude Tests: Biased Against You?" this past semester at Oglethorpe University in Atlanta and conducted a workshop on minimal competency testing for the Southeastern Teacher Corps Network.

Personal experiences with tests

As a child of urbane highly-educated middle-class parents, I personally have always done well on tests, scoring in the 99th percentile on the PSAT, SAT, MMSQT, and GRE. Therefore I was shocked to learn, while doing research in Mississippi, that Ole Miss's requirement (at the time) of an "800" on the GRE for admission to its theatre program meant, not a 99th percentile score, but two 400's (morning and afternoon sections of the test)! I thus realized the vast differences in performance by region. These differences relate to income and urbanity; family income alone correlates .83 with PSAT scores on a statewide basis, as shown in my article, "Breaking the Vicious Circle."

---

[1] Enclosed as Appendix A to this report is the Autumn-Winter issue of the "Clearinghouse for Civil Rights Research," which I edit. This issue, titled "Testing and Citizenship," includes four articles directly related to concerns of the Subcommittee.

[2] Enclosed as Appendix B is my vita.

Teaching at Tougaloo College provided further insights into the testing problems of Mississippians, this time black Mississippians. Tougaloo has a strong sociology program. Our graduates won ASA Minority Fellowships in each of the first two years of that program, Ford National Black Fellowships, Danforth Fellowships, and other national awards for graduate study. Yet our graduating seniors, all of whom were required to take the GRE, scored between 200 and 565 (the scale goes from 200 to 800 and beyond). We actually had students scoring 200, the minimum; a reasonably good student would receive perhaps 350, far below the national median, then 500.

Some students with scores in the 400's would nonetheless be admitted to difficult graduate programs (e.g., Princeton in international relations, Harvard in education and social policy, Washington University in sociology, Chicago in social work), usually because some member of the admissions committee knew Tougaloo's reputation in sociology and convinced the committee to disregard the GRE scores. (We also learned to help some students develop reasons why they had not taken the GRE, for some schools found it possible to waive the GRE requirement for a believable excuse, while they would not have overlooked a 400 score on it if taken.) In those graduate programs, our students did admirably, often completing advanced degrees in short order and with outstanding recommendations.

So I came to question the tests.

Now, let me add something that might perhaps weaken my case. The GRE did tend to put our students in the "correct" rank order. That is, our best majors—highest in grades and performance in our classes—tended to get the 460's and 530's. Our worst students received the 200's and 225's. There would be exceptions: a rather average white student at Tougaloo might get a 585, leading her entire cohort, while her actual academic performance was by no means that outstanding, while a particularly rural black student might receive a 270 while we regarded him much more highly. But in general the tests provided no real surprises, within our population of Southern black students.

When I taught at Harvard, I recall encountering one sophomore who could not do the work required by my relatively straightforward "Sophomore Seminar in Social Relations". Other students did not do the work: still others could do it only with difficulty, but this was the first Harvard student I had seen who, no matter how hard he worked, could not handle my course. So I checked his file. (This was in pre-Buckley Amendment days.) His SAT verbal aptitude was 485. He had graduated from Phillips Exeter Academy, his parents were highly educated, his father and grandfather were Harvard alumni, and he lived in an elite Northeastern suburb. Now I understood his problem. For although his 485 was identical to the 485 that might be scored by a strong sociology major at Tougaloo, among his peer group of urban prep-school students, it put him probably in the lowest quartile. He could not do the work!

And yet Educational Testing Service says, in its description of the GRE: The Aptitude Test, like other standardized tests, makes it possible to compare students with each other regardless of their individual backgrounds. A GRE score of 500, for example, has the same meaning whether earned by a student from a small private liberal arts college or by a student at a large public university. (1978–79 GRE Information Bulletin, p. 16)

This is not a factual statement, and as the next section will show, ETS has published no research to justify such a claim. It is simply a statement of arrogance!

My experiences in Mississippi allow an additional generalization. Because Millsaps College, a predominantly white institution in Jackson, had no strong methods/statistics course in sociology-anthropology, I was asked to develop one. For three semesters, I taught Methods and Statistics of Social Research simultaneously at a white and a black institution; during two of those semesters the instruction itself was partly merged, with Millsaps students coming to the Tougaloo campus for about a third of their classes. Assignments, texts, and tests were identical, and I graded examinations "blind," without knowing which students from which college had written them. Test scores were almost the same at the two schools, but a significant difference occurred in the composition of those scores. Tougaloo students did slightly better than Millsaps students on the extended essays or extended statistical problems, while Millsaps students definitely outscored their Tougaloo counterparts on the "short-answer" or "multi-choice" sections of the tests. These results, in addition to some close investigation of the performance of those Tougaloo students who did very poorly on multiple-choice questions, convinced me that persons deficient in language skills are particularly disadvantaged by multiple-choice testing.

This conclusion may seem contrary to common sense. Indeed, many of the very students afflicted by inadequate reading/writing skills clamored for multi-choice exams, thinking it would be easier for them to mark X's than to write paragraphs. Nevertheless, the process of reading a paragraph, holding it in mind while reading five sentences as alternatives A through E, testing each alternative against the paragraph, and selecting a best answer amounts to a particularly "white" and urban skill.[3] Thus the form of multiple-choice written exams discriminates against persons deficient in that skill.

This discrimination is justified to the extent that the skill involved is a fair representation of the abilities needed to do the job to which the examination deters entrance. However, the performance of black students in graduate school (and not just my own students, but those studied nationally, as references in the Appendix describe) is often excellent, despite their subaverage aptitude scores.

The same conclusion is justified by further analysis within my test results from Tougaloo and Millsaps. Black students did worse on the multiple-choice part of the exams. If those scores measured something basic, if they measured an important ability to deal with the subject matter of the course, then why did the same black students do slightly better than their white peers on essay parts of the exams?

The obvious overall conclusion from this experience, supported as well by other research,[4] is that the form of multiple-choice tests discriminates against minorities and non-urbane peoples, even when they have been exposed to the precise content on which they are being tested. (Bias in content is another matter; I treat it in the next section.) Other problems with form include the settings in which the tests are typically given ("white" in style; usually white in site and personnel as well) and the expectations of inferior minority performance (laid on black students with special severity by the society as a whole, by the educational system, and by their previous test experiences).

But if the form of the test is biased, that is a sweeping problem for the test-makers, and it cannot be dealt with by half-hearted measures. ETS, for instance, admonishes its test administrators to: Treat all candidates equally. Knowing how deeply minority candidates can be affected by the phychological atmosphere of the testing center, you will want to make certain that none of your procedures makes them feel different from other examinees. (ETS: Administering ETS Testing Programs, 1978–79: A Guide, 1978, page 20.)

Since the form of the test and of the testing does *not* "treat all candidates equally," this instruction is meaningless.

Nor are minorities the only persons adversely affected by bias in test form. Students at the University of Vermont fall into two basic categories: those from rural Vermont (perhaps excluding Burlington, Vermont's only sizable city) and those from suburbs of large Eastern cities. Median family income of the latter group is about $50,000; their SAT scores are about a hundred points above the average of "native Vermonters," as the phrase goes. Their freshman Grade Point Average is very slightly above that of Vermont students; by the senior year this small difference has reversed. It is possible, of course, that Vermont students work much harder to earn those basically equal grades, but I know of no evidence to support such a hypothesis. Again, the vast test score gap that exists across groups, in this case a rural/urban gap and to some extent an income gap, does not denote any commensurate differential in performance.

Research on test-score differentials

One reason I have spent so much time on my own experience with test-score differentials is that previous research in this area, at least as related to race and class, has been fairly scarce. Several problems afflict would-be researchers in this field. First, it is a sensitive issue. The College Board will not even allow ETS officially to release racial means, let alone make detailed data by race available, lest they be misinterpreted and misused by white supremacist writers.[5] Ironically, then, precisely because of this sensitivity to the issue, little research has been done on it, and some that has been performed seems more defensive than seeking in nature.[6] Research has also been inhibited by the testmakers' claimed need to keep their tests confidential and reusable, which restricts access to data and makes item-analysis almost impossible, at least by outsiders. Researchers within the testing agencies, on the other hand, tend to be limited to data from existing test items; this constricts the range of hypotheses they can investigate. Yet another obstacle is created by test-takers' need to focus on the test itself; hence they respond poorly to requests for socioeconomic data and academic performance information.

---

[3] See page 29 of Appendix A; cf. pages 6–11.

[4] See the studies by Abramson, Eckburg, Feldman and Sullivan, Haggard, Sattler, and Whimbey cited on pages 16–19 of Appendix A.

[5] George H. Hanford, Senior Vice President, The College Board, letter to Loewen, 11/10/78.

[6] "Some Points of Confusion in Discussing the Testing of Black Students," by ETS staffer Ronald L. Flaugher, begins with the statement, "There are several confusing issues that have delayed the progress toward seeing that testing is not a source of unfairness for minority students." (Chapter 2 of L. P. Miller, ed., The Testing of Black Students: A Symposium [Eaglewood Cliffs: Prentice Hall, 1974], 11.)

Within this bleak picture, there are some bright spots. A fairly large body of evidence links test performance with socioeconomic status and racial group memberships. (Of course, this evidence can itself be misused to "blame the victim" for poor performance.) Adequate documentation also indicates that test conditions typically favor majority students. Test anxiety, race of examiner, form of test, and other setting-related factors have been shown to depress minority students' scores disproportionately.

Bias in the content of tests has been claimed.[7] However, an ETS study of item bias, "Eliminating Differentially Difficult Items as an Approach to Test Bias," ETS Bulletin RB-78-4, found no such effect. These researchers followed a strict empirical approach. They removed those items from the SAT on which blacks did worst, vis-a-vis whites, in effect building new SAT verbal and math scales, and then they looked to see what had happened to scores. They found "relatively little effect." Hence they did not bother to look at the nature of the removed items. But their research was flawed by the enormous problem that the ratio of black success to white success on the removed verbal items was 58 percent (43.5 percent of the blacks got them right, compared to 75 percent of the whites), while on the retained items this ratio was 55 percent! Thus blacks did better on the removed items than on those items left in, vis-a-vis whites. No wonder their removal did not help blacks move up compared to the white median score! The researchers failed to do what they claim to have done.

Another ETS paper, "Word Associations of Students at Predominantly White and Predominantly Black Colleges," claims a different "null result." Joel T. Campbell and Leon H. Belcher gave a word association test, using words from the GRE, to students from black and white colleges. They indicate that the students did not take the test seriously, and they do not take it seriously either, in that they do not make any attempt to relate racially different word associations back to test performance; thus the words might just as well have been taken from any source in the English language. Despite these methodological inadequacies, black patterns were sometimes strikingly different from white responses, indicating that words are sometimes systematically misused or used differently in the black community. "Environment," for instance, was usually associated by black students with "home," "people," or "ecology," while whites associated it to "air," "clean," "ecology," and "earth." Neither usage is wrong; blacks are centering on the social environment while whites center on the natural environment. Obviously any syllogism, antonym, or other verbal ability item that used "environment" would have to be slanted, purposefully or accidentally, against one race or the other. Implications for verbal aptitude tests are substantial, especially where the word is merely used differently, not misused: items based on back usage could legitimately help "culture balance" other items based on vocabulary less known by blacks.

The foregoing analysis suggests that content bias can be subtle, as when common minority usage is different from majority meanings, or when Southerners think of one context for a word, Northerners and Westerners another. Much of the content bias in the GRE, SAT, Miller Analogies Test (MAT), LSAT, and other verbal aptitude tests is more blatent. Item after item is rather simple in reasoning or word relationships but is based on what can only be termed "establishment" vocabulary. I have never encountered an analogy based on joining terms, for instance, but have seen many based on Greek history or classical music. Items based on words like Thucydides, Herodutus, chianti, argentum, Hera, ambergris, Cather, cuneiform, Runnymede, minotaur, and Latinate test exposure, not aptitude, and America's residential and cultural segregation limits the exposure of minorities and rural children to this kind of information.

This is an international problem. According to the Organization for Economic Cooperation and Development, headquartered in Paris, in a study of testing and certification in Western developed nations generally: All selection mechanisms based on the usual indicators of education performance are invariably biased by factors related to the students' social and cultural background and thus necessarily help to maintain the existing social structure and its dominant values. (Selection and Certification in Education and Employment, 1977.)

Many European nations, once perhaps more rigid than the United States in their use of testing for tracking, have moved toward reform in this crucial area. Can we afford not to following their example?

---

[7] D. E. Coupland, "Aptitude Tests and Discrimination", International Law Review, Vol. 102, No. 3, 1970, 241-253.

Conclusion

1. "Standardized tests" are the greatest single barrier to equal opportunity for disadvantaged groups, at leat in the sphere of education. Minorities who get into higher education generally do so despite tests, not because of them, which is particularly ironic in light of the ostensible reason for the development of the tests themselves, to avoid petty prejudice and capricious or arbitrary rejections.

2. Many members of the majority, particularly persons from the working class or rural areas, are similarly disadvantaged. Test scores are drastically lower in some parts of the country than in others.

3. The vast test-score differentials across groups and regions are not validated by any commensurate performance differentials and therefore indicate a failure of the tests, not of the minorities or the regions.

4. A "truth in testing" bill is needed, among other reasons, so that social scientists outside the employ of the test companies can undertake unbiased research into the test-construction process, the effects of different kinds of items on different kinds of students, and the causes of differential performance across groups. Without this type of "sunshine legislation," it has not been easily possible to study these topics.

I thank you for your attention and urge your consideration of this legislation.

ADDITIONAL STATEMENT OF DR. JAMES W. LOEWEN, DIRECTOR OF RESEARCH, CENTER FOR NATIONAL POLICY REVIEW, SCHOOL OF LAW, CATHOLIC UNIVERSITY OF AMERICA, SEPTEMBER 24, 1979

In order to highlight two points about testing related to H.R. 3564 and H.R. 4949, and to add some specific comments about each bill, I ask that the Subcommittee allow me to add this statement to my statement of 8/15/79 and to its record in these hearings.

First, I submit additional data to emphasize the most basic issue testmakers have failed to address, therefore, the greatest problem that standardized tests pose to American society: tests may measure aptitude or achievement within populations that share backgrounds, but they do not measure accurately across backgrounds. Thus they discriminate against rural Americans, Afro-Americans, working-class students, and anyone else not from a white urban middle- or upper-class family. On this point, I enclose Table I, Aptitude and Per Capita Income by State. The aptitude data come from the PSAT of ETS, which the National Merit Scholarship Corporation has used as a part of their selection process for National Merit Finalists and Scholars. NMSC wants to foster a national competition, and in order to avoid giving an unfair advantage to residents of Princeton, suburban Connecticut, et al., it long ago institued different cutoff points for each state on the Preliminay Scholastic Aptitude Test/National Merit Scholarship Qualifying Test, cutoffs set so that the number of NMSC finalists is proportional to the number of high-school seniors. Connecticut has the highest cutoff, Mississippi the lowest, and it is intriguing to note that Connecticut is the richest state in the nation, Mississippi the poorest. As Table I indicates, 70% of all variation in PSAT cutoffs is associated with median income by state!

Incidentally, another variable, proportion urban in the population, plays a considerable role also. Its correlation with PSAT cutoffs, r=.57, explains 32.5 percent of the variation in PSAT cutoff scores.

Table I also shows by implication the problem a researcher faces in trying to learn about or explain regional, rural/urban, racial, or other variations in test scores: note tht I had to obtain these data from an "outside" source, NMSC; ETS would not release them to me; the better measure, state median score, therefore could not be studied.

Second, I enclose the "Loewen Low Aptitude Test" for your consideration. This teaching device is designed to show my urbane white students some of the forms of test bias and to give them the experience of "flunking" an aptitude test. Appendix C, a reprint of my article, "Introductory Sociology: Four Classroom Exercises," discusses the type of bias each question represents and provides (on page 239) the correct answers.

TABLE I.—APTITUDE AND PER CAPITA INCOME BY STATE

State	Income (1970)	Rank	PSAT cutoff [1]	Rank [2]
Alabama	$2,828	49	182	47
Alaska	4,676	4	194	9.5
Arizona	3,542	30	190	21.5
Arkansas	2,742	50	180	48.5
California	4,469	9	193	112.5
Colorado	3,751	21	191	19
Connecticut	4,807	2	200	1.5
Delaware	4,233	12	198	4.5
D.C.	5,519	1	200	1.5
Florida	3,584	29	190	21.5
Georgia	3,277	35	187	36.5
Hawaii	4,530	7	193	12.5
Idaho	3,206	39	188	34
Illinois	4,516	8	192	16
Indiana	3,773	20	189	27.5
Iowa	3,714	23	189	27.5
Kansas	3,804	18	189	27.5
Kentucky	3,060	43	183	45
Louisiana	3,065	42	180	48.5
Maine	3,243	37	189	27.5
Maryland	4,247	11	198	4.5
Massachusetts	4,294	10	197	7
Michigan	4,043	13	188	34
Minnesota	3,793	19	191	19
Mississippi	2,561	51	178	51
Missouri	3,659	26	189	27.5
Montana	3,381	34	189	27.5
Nebraska	3,700	24	189	27.5
Nevada	4,544	5	186	38.5
New Hampshire	3,608	27	191	19
New Jersey	4,539	6	198	4.5
New Mexico	3,044	45	188	34
New York	4,797	3	198	4.5
North Carolina	3,188	40	187	36.5
North Dakota	2,937	46	183	45
Ohio	3,983	15	189	27.5
Oklahoma	3,269	36	185	41
Oregon	3,700	24	192	16
Pennsylvania	3,893	17	193	12.5
Rhode Island	3,920	16	194	9.5
South Carolina	2,908	48	183	45
South Dakota	3,182	41	185	41
Tennessee	3,051	44	185	41
Texas	3,515	31	189	27.5
Utah	3,210	38	185	43
Vermont	3,491	32	192	16
Virginia	3,586	28	196	8
Washington	3,993	14	193	12.5
West Virginia	2,929	47	179	50
Wisconsin	3,722	22	189	27.5
Wyoming	3,420	33	186	38.5

r (Spearman rank-order correlation coefficient) = .83. r^2 = .689

[1] NMSC uses twice the verbal score (range 20 to 80) plus the math score (same range); the resulting scores range from 60 to 240.

[2] Ties have been assigned the mispoint of their ranking range.

r can vary from 0 to 1 in magnitude; $r=0$ means there is no relationship apparent between income and PSAT scores; $r=1.0$ means there is a perfect correspondence, so that knowing income the state PSAT cutoff can be predicted entirely without error. $r=1.0$ means there is a perfect correspondence, so that knowing income the state PSAT cutoff can be predicted entirely without error. $r=.83$ is a very high correlation coefficient. It means that almost 70% (68.9%) of all variation in state PSAT cutoff can be "explained" by one variable, median income.

Finally, my specific comments about each bill. As I read H.R. 3564, the "Truth in Testing Act of 1979," it requires nothing significant from, say, ETS, that is not now being provided. Therefore it seems paradoxical to me. Why enact it? If these hearings convince you there is no problem with standardized testing as now practiced in America, then I would think you would recommend no bill; if on the other hand these hearings convince you that there are problems that call for federal regulation, then you would want to recommend a bill that addresses some of the problems. H.R. 3564 does not.

Many of the problems with standardized testing cited in my previous statement and in this statement are considerably amelioriated by H.R. 4949, on the other hand. Its provisions regarding feedback of test questions and correct answers will help individuals and entire school systems learn areas in which they need improvement. Its provisions requiring filing of the relationships between test scores and income, race, etc., will help admissions officers understand that a 350 from rural Mississippi is not the same as a 350 from Scarsdale. Thus accountability and fairness will be increased.

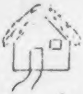
THE LOEWEN LOW APTITUDE TEST

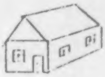
Directions: Answer every question. Do your best in the time allotted, for your entire future is at stake. Circle the letter corresponding to the single best answer to the question.

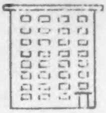
1. Nonverbal Picture Test

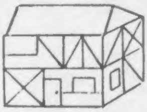
Hans and Bertha live in a thatched hut high in the Swiss Alps. There Hans herds her goats while Bertha goads his hurts. One day, after the sun came up . . .

Circle where Hans and Bertha live:


A.


B.


C.


D.

2-3. Analogies Test

2. Spline is to mitre as ______________ is to ________________.

a. love. . .marriage
b. straw. . .mud
c. key. . .lock
d. bond. . .bail
e. bond. . .paper

3. ___________ is to sake as opera is to ____________.

a. Pete's. . .many
b. hit. . .run
c. swordplay. . .Chianti
d. attend. . .forego
e. her. . .star

4-5. Verbal Aptitude and Reasoning Test.

Circle the letter which indicates the best answer.

4. Saturday Ajax got an I.D.

a. He had smoked too much grass
b. He tripped out on drugs
c. He brought her to his apartment
d. He showed it off to his fox
e. He became "wised up" (less dense)

5. Which of these situations is least congruous?

a. an Eames in the dining room
b. a Chippendale in the dining room
c. an Eames in the living room
d. a Chippendale in the living room
e. all of the above

Score	Est.SAT
0	200
1	200
2	275
3	400
4	525
5	700+

Correct ______ Estimated SAT______________ Est. Time_______________

I would make two minor suggestions regarding H.R. 4949. First, to the section requiring the report to Congress "concerning the relationship between the test scores of test subjects and income, race, sex, ethnic, and handicapped status," I would suggest you add "regional and rural/urban," for these are crucial variables inhibiting performance on the part of those who live in rural areas and in the South and parts of the West, Great Plains, and Midwest. (Sec. 4(b).

Second, why require provision to the Commissioner of "any study * * * for which [the test agency] provides data"? (Sec. 4.(a)(1)A)) This is too harsh on the test agencies, for it saddles them with a violation if they release data to a researcher over whom they then have no control, and who may not even send back to them a copy of his/her study. It might also become inadvertently counterproductive to the purposes of the bill, for it could deter test agencies from providing data to social scientists, lest they then unwittingly violate the law. The final phrase should be deleted, in my view.

Otherwise, I conclude from my knowledge of the use of standardized testing in American secondary and higher education, and from my attempts to do scientific research in this area, that H.R. 4949 would have a most salutary effect. A national industry is involved, and the federal regulations in this bill would be most helpful in eliminating some of the abuses now present in that industry.

Thank you for your consideration.

Mr. Weiss. the next witness is Ms. Simmons.

STATEMENT OF ALTHEA T. L. SIMMONS, DIRECTOR, WASHINGTON BUREAU OF THE NATIONAL ASSOCIATION FOR THE ADVANCEMENT OF COLORED PEOPLE

Ms. Simmons. Mr. Chairman and members of the subcommittee: I am Althea T. L. Simmons, director of the Washington Bureau of the National Association for the Advancement of Colored People. I am here today in behalf of the more than 1,800 branches, youth councils and college chapters of the NAACP which operate in the 50 States and the District of Columbia.

I am accompanied by Dr. Anderson J. Franklin, who is a member of the NAACP National Advisory Committee on Testing.

Professionally, Dr. Franklin is a professor of psychology and director of the Center for Academic Skills at the City College of the City University of New York.

I do not profess to be a psychometrician, that is why I have Dr. Franklin by my side who will be of assistance to me and the committee in responding to any technical question the committee may propound.

We appreciate the opportunity to appear before you to share the NAACP views on the important issue of testing and how it affects blacks and other minorities. I must share with the committee the grave concern the NAACP has regarding the wholesale use and misuse of standardized tests and the disproportionate impact such testing has on our constituents.

For 4 years I was the NAACP's national director for education programs and had occasion to work closely with all of our units on education issues, confer with many education officials at all levels and officials of the major testing companies on the consequences of ability grouping practiced by whole educational systems which results in racial isolation, the enforcement of stereotypes, the labeling of children and the reinforcement of feelings of inferiority which can lead and does lead to a third-class education.

From the cradle to the grave, there has been an increasing reliance in this country on the use of an assortment of standardized tests to screen, select, admit, reject, to classify, stratify, track, license or certify.

Such subjective dependence on objective tool makes it possible, and even attractive, in some quarters, to use tests inequalitably in ways which preclude a segment of the population from reaching the heights of its aspirations and realizing its potential.

The accelerated wide-scale use of standard tests limiting access of blacks and other minorities to institutions of higher education led to the mandate by the NAACP at its 65th Annual National Convention to its national office to call upon the Association of Black Psychologists and the College Entrance Examination Board to assert leadership in addressing themselves to the issue of standardized testing and the misuse thereof.

In 1972 the NAACP adopted a convention resolution calling for a moratorium on standardized tests until after other suitable and nonbiased criteria for measuring pupil progress and teaching accountability have been devised.

In line with the above statement of NAACP policy, I, in my then capacity as NAACP's national education director, convened a group of experts from the Association of Black Psychologists, the College Entrance Examination Board and the Educational Testing Service and the NAACP to begin dialog around this important issue.

Dr. Sidney Marland, president of the College Board, and vice president Dr. Stephen J. Wright and Dr. Belvin Williams, vice president of the Educational Testing Service were active participants and helped us to identify issues and conference participants. We convened a conference which produced the NAACP report on minority testing which is appended to this statement.

I cite this background to let the committee know the priority the NAACP gave to the impact of the unregulated use of standardized testing on blacks and other minorities.

The present legislation, targeted to postsecondary education, does not appear to take into account the use of standardized tests in the elementary and secondary schools.

We have spoken to legislators at the State level and with former Congressman Harrington who in the 95th session of Congress introduced H.R. 6776, a forerunner of the current legislation.

Although the Gibbons and Weiss bills are not as comprehensive as we would like, they are a step in the right direction. It is the position of the NAACP that truth-in-testing legislation is an idea whose time has come.

NAACP advocated such legislation in 1976 at our 67th Annual National Convention when a resolution was adopted calling for an Office of Consumer Affairs in Testing and Student Evaluation. That idea came from Dr. Franklin who accompanies me today. We felt there should be this kind of office set up on local and State levels where they would advocate adoption of a truth-in-testing law.

The NAACP is fully in accord with the purpose of the act as set forth in H.R. 4949. We believe that test publishers must be responsible for monitoring the use of their tests by assuming an advocacy role when necessary.

In our report on minority testing, supplied to the committee, we state that in instances where tests are used—or about to be used—

without due observance of this advocacy role, test publishers should apply sanctions, including the denial of the use of their product by those who misuse it.

Moreover, test publishers have an advocacy responsibility which requires that test objectives be clearly stated, that the process of administering the test be fully described so that subjects and users both understand exactly what is expected to happen and how it will happen, in nonthreatening, affirmative terms.

It is our understanding that there is little systematic information gathered on the relationship between test scores and future success, but indications are that this is a low correlation. The definition of future occupational success is ambiguous as many variables determine this which challenges the emphasis of test scores for admission.

The NAACP endorses test preparation courses so long as organizations and persons can verify their utility. Such organizations should also be regulated for their claims and services; likewise test agencies should be required to acknowledge and state to inquirers where coaching is effective.

We recognize that the test agency and tests are not wholly responsible for determining academic admissions or future performance; however, the test users could provide test agencies and public the estimated weight of percentage carried by tests in admissions consideration.

The NAACP is aware that test scores are being supplemented, in some instances, by other criteria, but we also know that college board scores are still the deciding factor regarding who gets admitted to what institution.

I could go into a list of horror stories on why this is necessary but time does not permit.

The widespread use of testing has already directed minorities into opportunities which will not allow them to perform well on postsecondary tests. The distinguished member of the New York Board of Regents, Dr. Kenneth Clark, said in his letter to the editor in the August 18, 1979, New York Times:

> Admission tests measure the abilities developed by individuals over a long period of time, both in and out of school. The value of test results is directly related to the equality of opportunity afforded to the candidates taking a particular test.

If tests used in elementary school already begin to put our children in a disadvantageous position, how can we reasonably expect them to be competitive on postsecondary education tests?

Teachers, counselors, and admissions officers are publics important to the testing industry since they sit at the entrance gates through which may test takers must go if they are to realize their goals.

The extent to which these publics can correctly interpret and effectively use those instruments is the extent to which decisions will be made that are fair to minority test takers.

The testing industry is called upon to help bridge that gap which exists between effective use and lack of understanding, by conducting workshops and institutes, and writing special publications that are aimed at interpretation of tests.

Mr. Chairman, we have a few recommendations to make to the committee, if we may. We believe that there is a need for clarity in the definition regarding who the test agency is so that the person who administers or develops tests knows the responsibility mandated under this bill.

Section 4(2)(b) is probably the best section in H.R. 4949 for it comes closer to what is needed in the whole area of testing. We believe there is a need for a mandate for the Commissioner to gather relevant data over a period of time to see what kind of things are going on. The data, I am reliably informed, is available in the computers, but there is feeling among professionals with whom I have conferred that they do not believe the testing companies will release such data unless there is a congressional mandate.

If the Commissioner has information on what is happening through the use of these tests, the Commissioner would then be in a position to report to the Congress and it could then make a determination if further action is needed.

We believe that it is necessary to establish a monitoring body for the collection of data from each testing agency. We believe the bill should spell out the type of data to be collected.

Under "Regulations and Enforcement," we were not certain whether the bill was calling for a civil penalty not to exceed $2,000 for each individual violation (one person) or group violation. In other words, if 10,000 persons are tested and there was a violation would the penalty be not more than $2,000 for the group of 10,000 or 10,000 times $2,000. This section needs clarification.

We believe that the bill dips into psychological jargon which will, in turn, lend itself to a response in the same vein. We would suggest that, wherever possible, the bill's language be simple enough for a layman, like myself, to understand.

We believe there needs to be massive data collected on what has happened to students who have been turned down because of their SAT scores or ACT scores.

As a former college teacher, I know of a number of instances where students were turned down by one school because of scores on standardized tests and were admitted to another institution and performed creditably.

It should be noted that this also pertains to preprofessional and postbaccalaureate such as LSAT and GRE tests. Test scores do not, we believe, predict precisely college or job success and they have a low relationship of what students are going to do academically when they get into school.

We want to diagnose what the students need and design a program to meet those needs. The intent of testing should be to open up avenues for students. Testing should be educational and not simply evaluative and discriminatory in nature.

ADDITIONAL RECOMMENDATIONS

I. The test user: (a) Test agencies be responsible for the training of test users. Test users should be informed of the validity, reliability and bias of tests. They should have at their disposal reports substantiating these factors as well as any other information which clarifies the use and limitations of tests.

Test users or any other contracting agency must demonstrate that they have personnel trained in the use and interpretation of tests.

Since misconceptions of tests can lead to damaging uses, the testing industry is obligated to make full disclosure of the purposes for which tests are designed, the processes by which they are designed, the population on which they were standardized, the statistical characteristics which delimit their use, for example, standard error of measurement, standard error of estimate, and other "do's and don'ts" which will affect optimal use and interpretation.

More specifically, it is incumbent upon the testing industry to give clear and precise interpretation of the scores their tests yield.

The public must be informed not only that the test scores are fallible and that their reliability is imperfect, but must be told the extent of that fallibility and imperfection. The public must be informed that test scores are only a sample of a student's performance and are never more than an estimate of truth.

CONSTRUCTIVE USES OF TESTS

Effective uses can be made of tests, regardless of their type, only if the user knows what the test contains, what its purposes are, and what its limitations are. We move on the assumption that an educational system, and particularly a school within the system, implicitly guarantees that students to whom achievement examinations—tests—are administered have been taught in ways such that they can reasonably be expected to have learned the information required by the examination.

Therefore, we call on the testing industry to assist school systems and other users in understanding better the content and constraints of the examinations, and in helping them understand how to make optimal use of the results.

Moreover, we call upon the educational community and the public private sectors to insure that tests are not used or relied upon where such guarantees do not exist.

Mr. Chairman and members of the committee, the NAACP strongly supports the bill H.R. 4949 because the key features provides for things never before available. True costs can be determined; real predictive validity can be approached on a job or school-related basis and more important, the monopoly would have public oversight.

[The complete statement of Ms. Simmons follows:]

PREPARED STATEMENT OF ALTHEA T. L. SIMMONS, DIRECTOR, WASHINGTON BUREAU OF THE NATIONAL ASSOCIATION FOR THE ADVANCEMENT OF COLORED PEOPLE, ACCOMPANIED BY DR. ANDERSON J. FRANKLIN, NAACP ADVISORY COMMITTEE ON TESTING AND PROFESSOR OF PSYCHOLOGY AND DIRECTOR, CENTER FOR ACADEMIC SKILLS, CITY COLLEGE OF CUNY

Mr. Chairman and members of the Subcommittee:

I am Althea T. L. Simmons, Director of the Washington Bureau of the National Association for the Advancement of Colored People. I am here today in behalf of the more than 1800 branches, youth councils and college chapters of the NAACP which operate in the 50 state ans the District of Columbia. I am accompanied by Dr. Anderson J. Franklin, who is a member of the NAACP National Advisory Committee on Testing. Professionally, Dr. Franklin is a professor of Psychology and Director of the Center For Academic Skills at the City College of the City University of New York. I do not profess to be a psychometrician, that is why I have Dr. Franklin by my side who will be of assistance to me and the Committee in responding to any technical question the committee may propound.

We appreciate the opportunity to appear before you to share the NAACP views on the important issue of testing and how it affects blacks and other minorities. I must share with the committee the grave concern the NAACP has regarding the wholesale use and misuse of standardized tests and the disproportionate impact, such testing has on our constituents.

For four years I was the NAACP's National Director for Education Programs and had occasion to work closely with all of our units on education issues, confer with many education officials at all levels and officials of the major testing companies on the consequences of ability grouping practiced by whole educational systems which results in racial isolation, the enforcement of stereotypes, the labeling of children and the reinforcement of feelings of inferiority which can lead to a third class education.

From the cradle to the grave, there has been an increasing reliance in this country on the use of an assortment of standardized tests to screen, select, admit, reject, to classify, stratify, track, license or certify. Such subjective dependence on "objective" tools makes it possible, and even attractive, in some quarters, to use tests inequalitably in ways which preclude a segment of the population from reaching the heights of its aspirations and realizing its potential. The accelerated widescale use of standard tests limiting access of blacks and other minorities to institutions of higher education led to the mandate by the NAACP's 65th Annual National Convention to its national office to call upon the Association of Black Psychologists and the College Entrance Examination Board to assert leadership in addressing themselves to the issue of standardized testing and the misuse thereof.

In 1972 the NAACP adopted a convention resolution calling for a moratorium on standardized tests until after other suitable and non-biased criteria for measuring pupil progress and teaching accountability have been devised.

In 1974 the following resolution on testing was adopted by the NAACP convention.

WHEREAS, a disproportionately large number of black students are being misplaced in special education classes and denied admission to higher educational opportunities.

WHEREAS, standardized tests, e.g., Stanford-Binet and the Wechsler Scale for Children, exclude blacks, Puerto Ricans and Mexican-Americans from the representative sample, and,

WHEREAS, such tests label black children as uneducable, assign them to lower educational tracks than whites; deny black children higher education opportunities; perpetuate inferior education; place black children in special classes and destroy growth and development of black children, and,

WHEREAS, students who fail to show a high verbal or numerical ability score low on the Scholastic Achievement Test (SAT), the Law School Admissions Test (LSAT), the Graduate Record Examination (GRE), etc., and are routinely excluded from college and graduate or professional education,

BE IT RESOLVED, that the NAACP demands a moratorium on standardized testing wherever such tests have not been corrected for cultural bias and directs its units to use all administrative and legal remedies to prevent the violation of students' constitutional rights through the misuse of tests, and,

BE IT FURTHER RESOLVED, that the NAACP calls upon the Association of Black Psychologists to assert leadership in aiding the College Entrance Examination Board and Educational Testing Service to develop standardized tests which have been corrected for cultural bias and which fairly measure the amount of knowledge retained by students regardless of his or her individual background,

BE IT FINALLY RESOLVED, that the NAACP directs its units to use all administrative remedies in the event of violation of students constitutional rights though the misuse of tests and directs National Office staff to use its influence to bring the CEEB, ET and ABP together to revise such tests.

In line with the above statement of NAACP policy, I, in my then capacity as NAACP's National Education Director, convened a group of experts from the Association of Black Psychologists, the College Entrance Examination Board and the Educational Testing Service and the NAACP to begin dialog around this important issue. Dr. Sidney Marland, President of the College Board and Vice President Dr. Stephen J. Wright and Dr. Belvin Williams, Vice President of the Educational Testing Service were active participants and helped us to identify issues and conference participants. We convened a conference which produced the NAACP Report on Minority Testing which is appended to this statement.

I cite this background to let the committee know the priority the NAACP gave to the impact of the unregulated use of standardized testing on blacks and other minorities.

The present legislation, targeted to postsecondary education does not appear to take into account the use of standardized tests in the elementary and secondary schools. We have spoken to legislators at the state level and with former Congressman Harrington who in the 95th session of Congress introduced H. R. 6776, a forerunner of the current legislation. Although the Gibbons and Weiss bills are not as comprehensive as we would like they are a step in the right direction. It is the position of the NAACP that truth in testing legislation is an idea whose time has come. NAACP advocated such legislation in 1976 at our 67th Annual National Convention when a resolution was adopted calling for an Office of Consumer Affairs in Testing and Student Evaluation:

OFFICE OF CONSUMER AFFAIRS IN TESTING AND STUDENT EVALUTION

WHEREAS blacks and other minorities are being increasingly subjected to various kinds of standardized tests which adversely affect their opportunities and accomplishments in later life, and

WHEREAS, there has been systematic exclusion of minorities at all phases of test development to utilization, and such exclusion presents a major concern in the evaluation of minority intellectual development and subsequent achievement in life, and

WHEREAS there is a need to halt current discriminatory testing practices,

NOW, THEREFORE, BE IT RESOLVED, that we direct our branches and state conferences to urge their respective states to establish a task force for the development of an Office of Consumer Affairs in Testing and Student Evaluation, or in the alternative assign this responsibility to an existing state agency capable of pursuing the objectives herein. The responsibilities of this agency would include, but not be limited to, the following:

(a) Consumer advocacy regarding the use and misuse of tests as well as advising parents of their legal rights in the testing of their children.

(b) Advocacy for the adoption of a "Truth-in-Testing" law.

(c) A test review board to scrutinize and systematically monitor test utility, development, policies and practices of all agencies employing assessment procedures.

(d) The development of informational advisory centers and layman documents on testing for parents of school children.

(e) The development of comprehensive Statewide Standards on Testing which reflect the interests of minority groups on the testing issue.

BE IT FURTHER RESOLVED, that we seek legislation in each state to establish a Research and Development Office which will have the latitude to study empirical questions of teacher and pupil performance in contrast to an office which functions as a statistics mill for legislative accountability.

The NAACP is fully in accord with the purpose of the Act as set forth in H. R. 4949. We believe that test publishers must be responsible for monitoring the use of their tests by assuming an advocacy role when necessary. In our Report on Minority Testing, supplied to the Committee, we state that in instances where tests are used (or about to be used) without due observance of this advocacy role, test publishers should apply sanctions, including the denial of the use of their product by those who misuse it.

Moreover, test publishers have an advocacy responsibility which requires that test objectives be clearly stated, that the process of administering the test be fully described so that subjects and users both understand exactly what is expected to happen and how it will happen, in non-threatening, affirmative terms. Publishers must state with clarity, in all descriptive information concerning a test they publish, the specific uses for which the test is designed, the specific limitations of the instrument and full explanation as to how the results should be interpreted.

§ 3. (a) (1). It is our understanding that there is little systematic information gathered on the relationship between test scores and future success, but indications are that this is a low correlation. The definition of future occupational success is ambiguous as many variables determine this which challenges the emphasis of test scores for admissions.

§ 3 (4) (C). The NAACP endorses test preparation courses so long as organizations and persons can verify their utility . Such organizations should also be regulated for their claims and services; likewise test agencies should be required to acknowledge and state to inquirers where coaching is effective.

§ 3 (5). We recognize that the test agency and tests are not wholly responsible for determining academic admissions or future performance; however, the test users could provide test agencies and public the estimated weight of percentage carried by tests in admissions consideration.

The NAACP is aware that test scores are being supplemented, in some instances, by other criteria, but we also know that College Board scores are still the deciding factor regarding who gets admitted to what institution.

May I take a few minutes to share with the committee a glaring example of how much weight is given to standardized test scores in a Virginia community several years ago. The local NAACP branch in Greenville County had gathered a large number of report cards of pupils who had been "partially promoted" as a result of a "get tough, learn or else stay behind" policy. A number of the students' report cards showed passing grades A's, B's and C's, yet the students were "partially promoted". Investigation on our part elicited information that the students whose cards we held had fared poorly on the standardized test administered by the school system. In a confrontation with school officials, school officials rejected the charge that standardized test scores were the sole criterion for the action; however, they were unable to explain how a student who recieved passing grades as evidenced by the report cards could be "partially promoted" if you disregarded the fact that they had scored poorly on the standardized test. The test scores were used to classify students into either college bound courses or OPT courses.

The widespread use of testing has already directed minorities into opportunities which will not allow them to perform well on post-secondary tests. The distinguished member of the New York Board of Regents, Dr. Kenneth Clark, said in his letter to the editor in the

August 18, 1979, New York Times:

> "Admission tests measure the abilities devleoped by individuals over a long period of time, both in and out of school. The value of test results is directly related to the equality of opportunity afforded to the candidates taking a particular test."

If tests used in elementary school already begin to put our children in a disadvantageous position, how can we reasonably expect them to be competitive on post-secondary education tests?

Teachers, counselors and admissions officers are publics important to the testing industry since they sit at the entrance gates through which many test takers must go if they are to realize their goals. The extent to which these publics can correctly interpret and effectively use those instruments is the extent to which decisions will be made that are fair to minority test takers. The testing industry is called upon to help bridge that gap which exists between effective use and lack of understanding, by conducting workshops and institutes, and writing special publications that are aimed at interpretation of tests.

PROMOTING A BETTER UNDERSTANDING OF TESTS

Sec. 5, 2 (a) (b)

Release of test questions and answers is an overstated concern. There are other professionals who will testify that it is not as technically serious as many tend to represent. It may increase cost somewhat but we believe the benefit gained by the individual, from learning of strengths and weaknesses in knowledge, far outweighs the monetary issue. Furthermore, there are reports that much of the claimed increased costs could be offset by using some of the already sizable profit margin enjoyed by ETS. We see the disclosure of test questions in the same spirit as teachers over the decades have returned corrected classroom tests, essays and term papers as a means of feedback on student performance. Likewise, we tend to compare the disclosure of test information to the public as similar to the need to know the effects of drugs on the human body. The life determining impact a test score can have on the future of an individual can be just as consequential as taking a drug. To date we have had no regulation or accountability for test agencies or their users which insures the consumer is getting a fair break.

Mr. Chairman, the NAACP believes that H. R. 3564 should call for a high degree of validity and reliability. It is our understanding from psychometricians that there is more than one way to compute a coefficient of reliability or a coefficient of validity and that it would not be amiss to have a company make known the formula used for such purpose since some formulae can give spurious, high coefficients.

The NAACP does not oppose the use of tests per se. We have called on the testing industry to develop standardized tools which have been corrected for cultural bias and which fairly measure the amount of knowledge retained by students regardless of his/her background. In 1974, NAACP called for a moratorium on standardized testing wherever such tests have not been corrected for cultural bias and directed its branches to use all administrative and legal remedies to prevent the violation of students' constitutional rights through the misuse of tests.

Ronald Samuda in Psychological Testing of American Minorities describes the problem we perceive thusly:

> "The system of higher education is inextricably interwoven with the total fabric of American society. Poverty, place of residence, the history of racial discrimination and segregation, the social stereotypes of race, the predominance of minorities within the lower stratum of the society, and the concomitant denial of educational opportunities - all these conspire to create and to perpetuate the imbalance and to entrench ethnic cleavages in the areas of higher education. The system, not a deliberate intrigue, perpetuates the boundaries that bar minorities from higher education and standardized testing plays a part (indeed quite a significant one)."

Mr. Chairman we have a few recommendations to make to the Committee, if we may. We believe that there is a need for clarity in the definition regarding who the "test agency" is so that the person who administers or develops tests knows the responsibility mandated under this bill.

Section 4 (2)(b) is probably the best section in H.R. 4949 for it comes closer to what is needed in the whole area of testing. We believe there is a need for a mandate for the Commissioner to gather relevant data over a period of time to see what kind of things are going on. The data, I am reliably informed is available in the computers, but there is feeling among professionals with whom I have conferred that they do not believe the testing companies will release such data unless there is a Congressional mandate. If the Commissioner has information on what is happening through the use of these tests, the Commissioner would then be in a position to report to the Congress and it could then make a determination if further action is needed.

We believe that it is necessary to establish a monitoring body for the collection of data from each testing agency. We believe the bill should spell out the type of data to be collected.

Under "Regulations and Enforcement", we were not certain whether the bill was calling for a civil penalty not to exceed $2000 for each individual violation (one person) or group violation. In other words, if 10,000 persons are tested and there was a violation would the penalty be not more than $2000 for the group of 10,000 or 10,000 times $2000. This section needs clarification.

We believe that the bill dips into psychological jargon which will, in turn, lend itself to a response in the same vein. We would suggest that, wherever possible, the bill's language be simple enough for a layman, like myself, to understand.

We believe there needs to be massive data collected on what has happened to students who have been turned down because of their SAT scores or ACT scores. As a former college teacher, I know of a number of instances where students were turned down by one school because of scores on standardized tests and were admitted to another institution and performed creditably. It should be noted that this also pertains to pre-professional and post-baccalaureate such as LSAT and GRE tests. Test scores do not, we believe, predict precisely college or job success and they have a low relationship of what students are going to do academically when they get into school.

We want to diagnose what the students need and design a program to meet those needs. The intent of testing should be to open up avenues for students. Testing should be educational and not simply evaluative and discriminatory in nature.

ADDITIONAL RECOMMENDATIONS

I. The Test User

(a) Test agencies be responsible for the training of test users. Test users should be informed of the validity reliability and bias of tests. They should have at their disposal reports substantiating these factors as well as any other information which clarifies the use and limitations of tests.

Test users or any other contracting agency must demonstrate that they have personnel trained in the use and interpretation of tests.

Since misconceptions of tests can lead to damaging uses, the testing industry is obligated to make full disclosure of the purposes for which tests are designed, the processes by which they were designed, the population on which they were standardized, the statistical characteristics which delimit their use, e.g., standard error of measurement, standard error of estimate, and other "do's and don't" which will affect optimal use and interpretation. More specifically it is incumbent upon the testing industry to give clear and precise interpretation of the scores their tests yield. The public must be informed not only that the test scores are fallible and that their reliability is imperfect, but must be told the extent of that fallability and imperfection. The public must be informed that test scores are only a sample of a student's performance and are never more than an estimate of truth.

CONSTRUCTIVE USES OF TESTS

EFfective uses can be made of tests, regardless to their type, only if the user knows what the test contains, what its purposes are, and what its limitations are. We move on the assumption that an educational system, and particularly a school within the system, implicitly guarantees that students to whom achievement examinations (tests) are administered have been taught in ways such that they can reasonably be expected to have learned the information required by the examination. Therefore, we call on the testing industry to assist school systems and other users in understanding better the content and constraints of the examinations, and in helping them understand how to make optimal use of the results. Moreover, we call upon the educational community and the public private sectors to ensure that tests are not used or relied upon where such guarantees do not exist.

[See also pages 22 and 29 (Minority Report)

II. Consumer Agency

See Minority Report.

Mr. Chairman and members of the Committee, the NAACP strongly supports the bill H.R. 4949 because the key features provides for things never before available. True costs can be determined; real predictive validity can be approached on a job or school-related basis and more important, the monopoly would have public oversight.

APPENDIX

RESOLUTIONS ON TESTING ADOPTED BY NAACP NATIONAL CONVENTION

YEAR	
1967	TESTING AND TRACK SYSTEMS
1968	TESTING AND TRACK SYSTEMS
1972	TESTING
1974	TESTING
1976	THE TESTING INDUSTRY
1976	FEDERAL REGULATORY AGENCY FOR TESTING
1976	NAACP REPORT ON MINORITY TESTING
1976	OFFICE OF CONSUMER AFFAIRS IN TESTING AND STUDENT EVALUATION
1976	MONITORING USE OF TESTS
1977	SPECIAL EDUCATION
1978	COMPETENCY-BASED TESTING / MINIMUM COMPETENCY TESTING
1979	NATIONAL TESTING MOVEMENT

OFFICE OF CONSUMER AFFAIRS IN TESTING AND STUDENT EVALUATION

WHEREAS, blacks and other minorities are being increasingly subjected to various kinds of standardized tests which adversely affect their opportunities and accomplishments in later life, and

WHEREAS, there has been systematic exclusion of minorities at all phases of test development to utilization,and such exclusion presents a major concern in the evaluation of minority intellectual development and subsequent achievement in life, and

WHEREAS, there is a need to halt current discriminatory testing practices,

NOW, THEREFORE, BE IT RESOLVED, that we direct our branches and state conferences to urge their respective states to establish a task force for the development of an Office of Consumer Affairs in Testing and Student Evaluation, or in the alternative assign this responsibility to an existing state agency capable of pursuing the objectives herein. The responsibilities of this agency would include, but not be limited to, the following:

(a) Consumer advocacy regarding the use and misuse of tests as well as advising parents of their legal rights in the testing of their children.

(b) Advocacy for the adoption of a "Truth-in-Testing" law.

(c) A test review board to scrutinize and systematically monitor test utility, development, policies and practices of all agencies employing assessment procedures.

(d) The development of informational advisory centers and layman documents on testing for parents of school children.

(e) The development of comprehensive Statewide Standards on Testing which reflect the interests of minority groups on the testing issue.

BE IT FURTHER RESOLVED, that we seek legislation in each state to establish a Research and Development Office which will have the latitude to study empirical questions of teacher and pupil performance in contrast to an office which functions as a statistics mill for legislative accountability.

ADOPTED 1976

BASIC POLICY STATEMENT
RESOLUTIONS ON EDUCATION
NATIONAL ASSOCIATION FOR
THE ADVANCEMENT OF COLORED
PEOPLE

THE TESTING INDUSTRY

WHEREAS, companies that develop, publish and sell tests should assume (or continue to assume) major responsibility for assuring the correction of deficiencies in their instruments, and

WHEREAS, the testing industry makes money from the administration of tests and has the responsibility of devising satisfactory and systematic ways of taking into account probable main effect-variables in the instructional setting.

THEREFORE, BE IT RESOLVED. that we call upon the testing industry, including the Educational Testing Service, American College Testing Program (ACT), Psychological Corp., Harcourt, Brace and Jovanovich, Science Research Associates, College Entrance Examination Board, CTB-McGraw Hill to state with clarity, on all descriptive information concerning a test, the specific uses for which the test is designed, the specific limitations of the instrument, and information as to how the results should be interpreted in acceptable professional practice;

BE IT FURTHER RESOLVED, that we urge the testing industry, at a minimum to include within the information it publishes concerning standardized tests of ability, achievement, personality and any other assessment procedures, specific data regarding predictive content and prescriptive validity;

BE IT FURTHER RESOLVED, that we call upon the testing industry to establish and fund an independent research and development corporation charged with the responsibility to (1) identify the critical problems in assessment as they relate to minority groups, (2) to sponsor research to investigate those problems requiring study, (3) to sponsor appropriate development work and to involve researchers who have the endorsement of minority professional and community associations and to set aside a minimum of 4 percent of net income to support these objectives;

BE IT FURTHER RESOLVED, that we urge the testing industry to convene a group of national black organizations to initiate dialogue around the implementation of recommendations and issues raised in this resolution.

BE IT FINALLY RESOLVED, that we call upon the National Office, through the Education Department, to assist local units in resolving problems arising from the misuse of tests in their locality.

ADOPTED 1976
BASIC POLICY STATEMENT
RESOLUTIONS ON EDUCATION
NATIONAL ASSOCIATION FOR
THE ADVANCEMENT OF COLORED
PEOPLE

FEDERAL REGULATORY AGENCY FOR TESTING

WHEREAS, the Congress through recent Federal legislation is placing increasing emphasis on program evaluation with the recent amendment of Title I of the Elementary and Secondary Education Act (as amended by PL 93-380) charging the Commissioner of Education with the responsibility of developing and publishing "Standards and Models for Program Evaluation," including the use of specific techniques and methodologies for such evaluations, and

WHEREAS, the Office of Education, in carrying out this Congressional mandate, is publishing a Series of Monographs on Evaluation in Education covering various issues regarding testing and evaluation, and

WHEREAS, the testing industry is a multi-million dollar industry, receiving millions of Federal dollars, and

WHEREAS, Title IV of the Elementary and Secondary Education Act (ESEA) as amended gives the state increased responsibility in the area of program evaluation (including program evaluation) and states are currently selecting advisory councils and developing methods of evaluation, and

WHEREAS, the Education of the Handicapped Act of 1975, recognized the importance of protecting children from misdiagnosis by insisting on the development and use of "culture fair tests" and the bill (PL 94-142) includes a short statement of the importance of gathering data from various sources rather than a single test, when diagnosing a handicapped condition, and

WHEREAS, the Developmentally Disabled Assistance and Bill of Rights (PL 94-103) calls for the establishment of a comprehensive system for evaluating the progress of individuals with developmental disabilities,

NOW, THEREFORE, BE IT RESOLVED, that we call upon the Congress to establish a Federal Regulatory Agency for the testing industry similar to the Federal Communications Commission, the Food and Drug Administration and the Environmental Protection Agency for the protection of the citizen consumer of tests and evaluative and assessment procedures.

ADOPTED 1976
BASIC POLICY STATEMENT
RESOLUTIONS ON EDUCATION
NATIONAL ASSOCIATION FOR
THE ADVANCEMENT OF COLORED
PEOPLE

MONITORING USE OF TESTS

WHEREAS, standardized testing is increasingly used to classify stratify and certify individuals and groups in our society, and

WHEREAS, the determination of public policy related to the testing of minorities has been insidious and extremely deleterious, and

WHEREAS, there is a need for more fair and effective processes of assessment which more adequately consider the diverse abilities, talents and cultural backgrounds of children if standardized tests are to be used, and

WHEREAS, test scores influence the way children think about themselves; and teacher expectations of the children, and

WHEREAS, there is increasing evidence that tests, as used, are determining what is being taught, rather than measuring what is being learned.

BE IT RESOLVED, that NAACP units mount a campaign against misuse of tests by urging parents to question not only their own child's test scores but also the testing programs in their school, city and school district.

BE IT FURTHER RESOLVED, that the NAACP units, through their education committees, make inquiries of testing policies and practices in their school districts, regarding:

1) What standardized IQ and achievement tests are given in the school district? at what ages?

2) How, and by whom, are the tests selected? Can the tests be seen and reviewed? By whom and under what conditions?

3) What learning goals do achievement tests serve? What is the impact of the various tests used on what is taught in the curriculum, on books and other materials purchased?

4) What decisions are made on the basis of test scores (e.g., placement in "tracks," special education, recommendation to college)?

5) Are tests used in the local school district free of ethnic, racial and class bias? Are they fair or unfair to children of different backgrounds?

BE IT FURTHER RESOLVED, that we direct our branches to request copies of all tests currently used by the local school district as a basis for educational decisions regarding pupil placement, assignment, promotion or academic programming.

BE IT FINALLY RESOLVED, that the results of such survey and copies of tests secured be forwarded to the Education Department of the National Office for follow-up, including the Education Department's calling upon the Department of HEW to conduct systematic analyses of the federally-funded testing and assessment programs or instruments used in public schools across the country.

ADOPTED 1976

BASIC POLICY STATEMENT
RESOLUTIONS ON EDUCATION
NATIONAL ASSOCIATION FOR
THE ADVANCEMENT OF COLORED
PEOPLE

TESTING AND TRACK SYSTEMS

The practice of placing children in "ability groupings" based upon culturally biased tests has been discredited by leading educational authorities. As a result of these testing procedures, many Negro students are placed at an initial disadvantage and are relegated to slow learning tracks from which they rarely escape.

The discrimination aspects of the race and class concepts underlying the tracking system were starkly set forth by Judge J. Skelly Wright in the Hobson V. Hansen decision in which the tracking system in Washington, D. C., was declared unconstitutional.

We call on NAACP branches to vigorously oppose homogeneous groupings or tracking systems wherever they discriminate in this manner.

ADOPTED 1968
BASIC POLICY STATEMENT
RESOLUTIONS ON EDUCATION
NATIONAL ASSOCIATION FOR
THE ADVANCEMENT OF COLORED
PEOPLE

TESTING AND TRACK SYSTEMS

The concept of placing children by means of "ability grouping" based upon the results of tests administered in the early grades has been discredited by educational experts and authorities. Virtually all tests used are culturally biased and unsuitable for actually determining the ability to learn. As a result of these testing procedures, many Negro students are placed at an initial disadvantage and are relegated to slow-learning tracks from which they rarely escape. As a result, the overwhelming proportion of Negro students are thereby exposed to a subtle, but systematic, form of discrimination within the school and class which does not prepare them for other than menial jobs. Moreover, the use of these devices tends to stigmatize the Negro student as intellectually inferior.

ADOPTED 1967
BASIC POLICY STATEMENT
RESOLUTIONS ON EDUCATION
NATIONAL ASSOCIATION FOR
THE ADVANCEMENT OF COLORED
PEOPLE

NAACP REPORT ON MINORITY TESTING

WHEREAS, the NAACP expressed its concern re the disproportionately large number of black students being misplaced in special education classes as a result of scores on standardized tests, and

WHEREAS, the 1974 convention mandated the Education Department at the National Office to use its influence to bring together the Educational Testing Service, the College Entrance Examination Board and the Association of Black Psychologists regarding this issue, and

WHEREAS, the Education Department has published the NAACP REPORT ON MINORITY TESTING, setting forth recommendations and an action agenda;

NOW, THEREFORE, BE IT RESOLVED, that we direct our units, youth and adult, to implement the recommendations of the Report.

BE IT FURTHER RESOLVED, that the NAACP Education Department establish a national task force to develop specific guidelines for laymen's participation in and support of standardized assessment procedures and disseminate the guidelines to the broadest possible audience.

BE IT ALSO RESOLVED, that we direct our units to establish contact with local chapters of the Association of Black Psychologists for technical assistance in implementing this resolution.

ADOPTED 1976
BASIC POLICY STATEMENT
RESOLUTIONS ON EDUCATION
NATIONAL ASSOCIATION FOR
THE ADVANCEMENT OF COLORED
PEOPLE

TESTING

WHEREAS, the National Black Psychologists have called for a moratorium on standardized testing; and,

WHEREAS, the 1972 National Education Association conferees recommended that IQ scores be removed from pupil/student cumulative record folders; that the National Teachers Examination be abolished; that NEA provide legal aid to persons suing against misuse of tests; and that colleges and universities eliminate the use of the Graduate Record Examination and Millers Analogy; and,

WHEREAS, there is evidence to suggest that intelligence tests do not represent a true measure of the scholastic experience or ability of the child, be it hereby

RESOLVED, that all units of the NAACP call upon their school administrators to be alert to the forthcoming recommendations of the National Black Psychologists Association and be prepared to declare a moratorium on standardized testing after other suitable and non-biased criteria for measuring pupil progress and teaching accountability has been devised.

ADOPTED 1972

BASIC POLICY STATEMENT
RESOLUTIONS ON EDUCATION
NATIONAL ASSOCIATION FOR
THE ADVANCEMENT OF COLORED
PEOPLE

TESTING

WHEREAS, a disproportionately large number of black students are being misplaced in special education classes and denied admissions to higher educational opportunities.

WHEREAS, standardized tests, e.g., Stanford-Binet and the Wechsler Scale for Children exclude blacks, Puerto Ricans and Mexican Americans from the representative sample, and,

WHEREAS, such tests label black children as uneducable, assign them to lower educational tracks than whites; deny black children higher education opportunities; perpetuate inferior education; place black children in special classes and destroy growth and development of black children, and,

WHEREAS, students who fail to show a high verbal or numerical ability, score low on the Scholastic Achievement Test (SAT), the Law School Admissions Tests (LSAT), the Graduate Record Examination (GRE), etc., and are routinely excluded from College and graduate or professional education,

BE IT RESOLVED, that the NAACP demand a moratorium on standardized testing wherever such tests have not been corrected for cultural bias and direct its units to use all administrative and legal remedies to prevent the violation of students' constitutional rights through the misuse of tests, and

BE IT FURTHER RESOLVED, that the NAACP call upon the Association of Black Psychologists to assert leadership in aiding the College Entrance Examination Board to develop standardized tests which have been corrected for cultural bias and which fairly measure the amount of knowledge retained by students regardless of his or her individual background.

BE IT FINALLY RESOLVED, that the NAACP direct its units to use all administrative remedies in the event of violation of students constitutional rights through the misuse of tests and directs National Office staff to use its influence to bring the CEEB and ABP together to revise such tests.

ADOPTED 1974
BASIC POLICY STATEMENT
RESOLUTIONS ON EDUCATION
NATIONAL ASSOCIATION FOR
THE ADVANCEMENT OF COLORED
PEOPLE

NAACP REPORT ON MINORITY TESTING

Published by the
NAACP Special Contribution Fund

FOREWORD

Some people seem to think that "intelligence tests" measure intelligence. I've got news for them. "Intelligence tests" do not measure intelligence: they measure one's ability to react to intelligence tests.

I know a man who was tested to determine his dexterity and quickness in seeing relationships between shapes and forms--a very important part of intelligence. On that test, he came out a little lower than a low-grade moron. Then he took a second test, to measure another important component of intelligence, accuracy of memory and quickness of recall. This time, he came out a little higher than a high-grade genius. Was he a moron or a genius?

I once watched two persons taking the same test for accuracy of memory and quickness of recall. They showed startling differences: one ranked very high, the other quite low. The reason was not hard to find: they came from opposite sides of the tracks. The test items were things which had been part of the intimate daily life of one, while the other had never seen and seldom heard of the items mentioned in the test.

Another case in point, directly relevant to the question of alleged "racial" difference in intelligence, was the administration of a test to children in Central Harlem a few summers ago. The children were told to indicate whether a lark was a dog, an automobile, a bird, or a kind of cheese ("Choose one.") None of them answered the question. Mindful of the "LARK" plastered throughout the neighborhood on giant-sized billboards, they later explained their failure on the test question by protesting, "It didn't say 'cigarette'!"

If the matter stopped there, it would be serious enough; but it does not stop there. Not only are individual children tested in order to be detested, standardized tests of intelligence are being used throughout the nation, in ways which permanently affect the self-image of the tested, thereby profoundly altering their own expectations as to their probable futures--and equally affecting the expectations of their teachers. Whole schools, entire school systems, states and regions, are compared to each other on the basis of uniform standardized tests. Racial and ethnic groups are stigmatized. The alleged differences thus "scientifically" validated become self-fulfilling prophecies.

But that's not all. The general mind-set of the nation welcomes these results, because, by training and experience, most people have a need to have somebody else they can feel "better than." We are a highly competitive collection of peoples, living in a nation where competitive sports outdraw all other television audiences. The desire to win is pandemic. From earliest childhood, we are praised for winning--sometimes consoled and often reprimanded for losing. The almost psychotic fervor of a Little Leaguer's parents dominates our culture. And since the schools are part of that culture, each individual is encouraged to out-do his peers. Graded on a bell curve, tested and compared to averages and norms, a child is rewarded or punished by the educational system on the basis of his success in outdistancing others.

"So what's wrong with that!" the true-believing American demands. Much is wrong with it. Of course there is nothing wrong--indeed much is right--in stressing "excellence." But we don't stress "excellence;" we stress "excelling." To be excellent is to shine: to excel is to out-shine. In our national effort to be Number One, we miss the goal of excellence. We settle for its cheap and deceitful substitute, excelling.

That is why a Parent-Teacher Association the other night complained bitterly to a bewildered principal that in every grade half of the children were below the median of their grade. Unfortunately, the general understanding of the meaning of standardized tests is about on the level of that PTA.

The necessity, compulsion, drive to excel others vitiates the more admirable principle of becoming excellent. And because the general public has much at stake in this matter of serving the purposes of a put-down society, the materials in this little Report are important. One doubts that, as a nation, we will quickly put aside our boastful competitive adversary urge to excel; but we ought, at the very least, to observe the elementary dictates of good sportsmanship and fair play. We ought no longer to use standardized tests to announce to the world that some children are "better" than others--particularly when elements of unfairness are built into the testing programs. (Remember the "lark" question for city children.) But even if the tests could be made "culture-free," their use would still be questionable because they are used to predict a child's future, condemning some to be "slow learners"--for life.

This nation ought no longer to permit pseudo-scientists to tell us that some racial groups are inferior to others. We ought not to stand quietly by, permitting the results of nation-wide testing to be used to destroy the self-image of some while falsely inflating the self-image of others.

The men and women who have invested their professional careers in the testing industry are not knaves or fools; but neither are they saints. We should not despair of them. Like all the rest of us, they should be presumed to have the ability to learn. Among other things, this means that we must presume that, given sufficient stimulus, they will put off their old ways and begin to direct their industry toward the production of a new and acceptable product; a testing system designed not to measure degrees of excelling in a competitive rat race but to promote self-understanding in all, as each becomes more excellent. We have no business permitting schools to be run as though life were a rat race. We should be interested in but one race, the human race.

With something of this purpose in mind, the Conference on Testing was convened. The membership of the conference was representative of many points of view. Noticeably lacking, however, was representation from the elitist establishment whose anti-democratic bias makes them perfect practitioners of the art of the put-down. Another unfortunate hiatus in the conference membership was occasioned by the absence of representatives from the two largest customers of the testing industry, the Federal and State Educational

Establishments. Never mind. These others have been having their say and they do not lack a forum. It is our belief that the time has come for another voice to be heard, a voice which speaks without equivocation in behalf of equity and decency and fair play and mutual respect: a voice which will not cease as long as inequity, indecency, unfairness and mutual contempt are thrust upon the lives of our children by elders who should know better.

•

May I thank the participants in the Conference, even as I apologize for what I have done to their carefully crafted reports from the four Task Forces. In editing, I have endeavored to retain as much as I could of the technical professional language which is natural to them as they communicate with each other; but I have also been mindful of the needs of the general reader. I may have sacrificed something of technical accuracy in order to achieve lucidity. If experts in the field of testing find the language sometimes less than completely professional, they must put the blame not on a supposed incompetence of the Conference participants but upon the tender concerns of an editor who wanted common people like himself to understand what was being said.

BUELL G. GALLAGHER

Vice Chairman, Emeritus,
National Board of Directors,
National Association for the
Advancement of Colored People

May 17, 1976

PARTICIPANTS IN NAACP INVITATIONAL CONFERENCE ON MINORITY TESTING

Dr. William F. Brazziel, Professor of Higher Education, University of Connecticut

Mr. Silas Craft, NAACP National Board of Directors

***Dr. Anderson J. Franklin, Associate Professor, The City College of the City University of New York and the Rockefeller University

Dr. William Hayes, Chairman, Association of Black Psychologists

*Dr. Asa Hilliard, Dean, School of Education, San Francisco State University

Dr. George D. Jackson, Professor School of Social Work, Howard University

**Dr. Walter J. Leonard, Assistant to the President, Harvard University

Mrs. Thelma C. Lennon, Director, Division of Pupil Personnel Services, Department of Public Instruction, State of North Carolina

Dr. Arthur L. Mathis, Director, Research and Evaluation, Community Mental Health Center, Meharry Medical College

Dr. Ann Borders-Patterson, University of North Carolina at Chapel Hill

Mr. Silas Purnell, Ada S. McKinley Community Center, Chicago

***Dr. S. O. Roberts, Chairman, Department of Psychology, Fisk University

****Dr. Melvin P. Sikes, Professor of Educational Psychology, University of Texas at Austin

Dr. Madelon D. Stent, Professor of Education, The City College of the City University of New York

RESOURCES CONSULTANTS

Mr. Lawrence Barclay, Minority Affairs Officer, College Entrance Examination Board

Dr. Edmund Gordon, Director, Institute for Urban and Minority Education, Educational Testing Service

George H. Hanford, Senior Vice President, College Entrance Examination Board

Michael Meyers, Assistant Director, NAACP

Dr. E. Belvin Williams, Vice President, College Board Programs, Educational Testing Service

Julius E. Williams, Director, Armed Services & Veterans Affairs, NAACP

*****Althea T. L. Simmons, Director for Education Programs, NAACP

Dr. Stephen J. Wright, Vice President, College Entrance Examination Board

RECORDERS

Ms. Ana Aponte, Secretary, NAACP

Benjamin J. Colbert, Assistant Program Director, Educational Testing Service

William U. Harris, Area Director, Admissions and Guidance, Educational Testing Service

Ms. Yvonne Ladson, Assistant to the Director for Housing Programs, NAACP

Ms. Esther Spann, Secretary, NAACP

---

*Chairman, Task Force on Use and Misuse of Tests
**Chairman, Task Force on Code on Testing
***Chairman, Task Force on Psychometric Integrity of Tests
***Co-Chairman, Task Force on Psychometric Integrity of Tests
****Chairman, Task Force on Social Policy
*****Conference Convener

DISCLAIMER BY THE COLLEGE ENTRANCE EXAMINATION BOARD

The College Entrance Examination Board has participated actively in, and provided partial financial support for, the deliberations which have led to the publication of this report. College Board participation has been with the firm belief that vigorous and continuing efforts should be made to improve standardized tests, to see that they are sensitively administered and accurately interpreted, and to eliminate testing abuses of all kinds. There are, however, several recommendations in the report, including the one pertaining to a moratorium which the College Board cannot support.

Nevertheless, the College Board will give the most serious attention to those recommendations having to do with the improvement of tests, their administration and their interpretation. Moreover, it will continue to be particularly vigilant with respect to any and all abuses.

DISCLAIMER BY THE EDUCATIONAL TESTING SERVICE

The Educational Testing Service has participated actively in, and provided partial financial support for, the deliberations which have led to the publication of this report. Educational Testing Service has participated in the firm belief that vigorous and continuing efforts should be made to improve standardized tests, to see that they are sensitively administered and accurately interpreted and to eliminate testing abuses of all kinds. There are, however, several recommendations in the report, including the one pertaining to a moratorium, which it cannot support.

On the other hand, the Educational Testing Service will give the most serious attention to those recommendations having to do with the improvement of tests, their administration and their interpretation. Moreover, it will be particularly vigilant with respect to any and all abuses.

CONVENER'S INTRODUCTION

The Conference on Minority Testing held in September, 1975 was convened by the National Association for the Advancement of Colored People. It reflects the NAACP's concern with the negative impact on Blacks of the national trend toward assessing aptitude, achievement and intelligence almost solely through test instruments, the validity and/or administering of which is often suspect. This trend which began with the mass objective testing during World War I, has accelerated to the point where in the last two decades great emphasis has been placed on group assessment, which often appear to operate to the detriment of many Blacks.

The assessing of achievement and intelligence of Blacks has been a concern to the NAACP since its founding, for the theme of the NAACP Founding Conference in 1909, was a refutation of arguments that the Negro was physically and mentally inferior. At that historic conference, leading scientists presented papers to refute that widely-held opinion.

During the ensuing 66 years, the issue has been periodically debated, either formally or informally, by NAACP and others, from time to time. The writings of the Moynihans Jensens and Schockleys and the key Supreme Court decisions in Hobson v. Hansen, Griggs v. Duke Power Co. and DeFunis v. Odegaard have stirred much controversy. More and more, tests have been utilized as the sole, or principal, means of making assessments under circumstances which may adversely influence one's opportunities and achievements throughout one's entire life.

The NAACP has been particularly concerned with the consequences of ability grouping practiced by whole educational systems which results in racial isolation, the enforcement of stereotypes, the labeling of children, and in the reinforcements of feelings of inferiority which can lead to a third class education.

The following vignettes of experiences related to our more than 2,000 units across the country illustrate the scope of the problem:

- Examination scores are used to determine who has access to so-called "examination schools" and who is admitted to the more prestigious colleges and universities.
- Classification systems based on standardized tests label a disproportionately large number of minority children as subnormal and a disproportionately small number of black and other minority children as gifted.
- Many black teachers are kept out of the classroom based on their scores on the National Teachers Examination.
- Many minority group students are unable to enter graduate and/or professional school because of test scores.
- Students are placed in EMR or other special education classes on the basis of test scores.

In 1972 the NAACP adopted a convention resolution calling for a moratorium on standardized tests after other suitable and non-biased criteria for measuring pupil progress and teaching accountability has been devised.

In 1974 the following resolution on testing was adopted by the NAACP convention.

WHEREAS, a disproportionately large number of black students are being misplaced in special education classes and denied admissions to higher educational opportunities,

WHEREAS, standardized tests, e.g., Stanford-Binet and the Wechsler Scale for Children exclude blacks, Puerto Ricans and Mexican-Americans from the representative sample, and,

WHEREAS, such tests label black children as uneducable, assign them to lower educational tracks than whites; deny black children higher education opportunities; perpetuate inferior education; place black children in special classes and destroy growth and development of black children, and,

WHEREAS, students who fail to show a high verbal or numerical ability, score low on the Scholastic Achievement Test (SAT), the Law School Admissions Test (LSAT), the Graduate Record Examination (GRE), etc., and are routinely excluded from college and graduate or professional education,

BE IT RESOLVED, that the NAACP demand a moratorium on standardized testing wherever such tests have not been corrected for cultural bias and direct its units to use all administrative and legal remedies to prevent the violation of students' constitutional rights through the misuse of tests, and,

BE IT FURTHER RESOLVED, that the NAACP calls upon the Association of Black Psychologists to assert leadership in aiding the College Entrance Examination Board and Educational Testing Service to develop standardized tests which have been corrected for cultural bias and which fairly measure the amount of knowledge retained by students regardless of his or her individual background,

BE IT FINALLY RESOLVED, that the NAACP directs its units to use all administrative remedies in the event of violation of students constitutional rights through the misuse of tests and directs National Office staff to use its influence to bring the CEEB, ETS and ABP together to revise such tests.

Following the adoption of this resolution, the Education Department of the NAACP extended invitations to the Association for Black Psychologists, The College Entrance Examination Board and the Educational Testing Service to meet with us to discuss the concerns of our Association regarding testing. During the ensuing year, representatives from these organizations met with us on three occasions to identify and focus on selected key issues. The results of our deliberations led us to convene the Conference on Minority Testing.

This Conference was not designed to resolve all the issues in testing. Rather it was designed to explore certain issues regarding how testing impacts public policy; whether or not there should be a code on testing and if so, what should the code encompass; the use and misuse of tests and the psychometric integrity of tests. The specific objectives as outlined to the conferees, was to elicit:

1. a set of specific recommendations that seek to deal with the issues and problems that have been identified.
2. a rationale for the Task Force's recommendations that give meaning.

to and a basis for interpreting the recommendations.

3. suggestions for ways and means to implement the recommendations.
4. a summary of all issues considered with pros and cons.

The relevance of the issues addressed during the Minority Testing Conference was underscored in many ways. To mention two, we can point to the fact that participants at the Conference were educators, representatives from the testing industry, professional and community organizations and, secondly, we noted the many common threads running through the four Task Force reports.

Those threads include an awareness that:

- Some type of assessment is needed.
- Tests, in our credential-oriented society, vastly influence the economic potential of human beings.
- There is a need for widepsread dissemination of information regarding testing in a form easily understood by all segments of the population.
- Test developers have an advocacy role to perform including sanctions for continued abuses.
- Test developers have a responsibility to tell what tests do and do not measure.
- Persons who take standardized tests must know their rights.
- Guidelines for the administration of tests should include specifics regarding the type of environment necessary for optimum test performance.
- Subjects and users must understand what is expected.
- Criterion-referenced approaches and materials should be further investigated.

In view of the fact that tests, (since they stratify or certify people) are, in effect, married to national policy issues, and determine what kinds of people, from what background, where they will fit in the society, what role they will play, and the like, it is incumbent on the Minority Community to galvanize educators, the testing industry, parents, students and community organizations to work systematically to insure that the assessment of individuals is culturally fair.

This is vital if all Americans are to enjoy equal opportunity in every aspect of public life.

The NAACP expresses its sincere appreciation to the Task Force Chairmen and the support staff.

We are also grateful to the individuals and organizations that contributed materials and data as background information and to the College Entrance Examination Board and the Educational Testing Service for the financial assistance which made this Conference possible.

ALTHEA T. L. SIMMONS

Director for Education Programs
The National Association for the
Advancement of Colored People.

May, 1976

SUMMARY OF RECOMMENDATIONS

1. That there be a moratorium on all current standardized tests unless such instruments conform to recommendations set forth in this Report.

2. That a national monitoring body be established with the power to enforce, through sanctions, to assure proper assessment and policy regarding the administration of assessment tools.

3. That companies that develop, publish and sell tests assume (or continue to assume) major responsibility for assuring the correction of the deficiencies in their instruments.

4. That the testing industry be held responsible for the development of assessment procedures which conform to professional standards as described in Standards for Educational and Psychological Tests developed by the Joint Committee of the American Psychological Association, American Educational Research Association and the National Council on Measurement in Education.

5. That the testing industry, at a minimum, include within the information it publishes concerning standardized tests of ability, achievement, personality and any other assessment procedure, specific data regarding predictive, content and prescriptive validity.

6. That, where standardized assessment results in the disproportionate sorting of groups according to ethnicity, the test developer provide separate validity coefficients for ethnic groups to which the assessment procedure is to be applied.

7. That test developers describe the probable main effect-variables in the instructional setting, in standardized terms, which must be considered along with the results from standardized testing if the interpretations of results are to be meaningful or acceptable. For example, the ethnic background of the examiner is known, in some cases, to affect the scores which children earn on IQ tests; therefore information about the ethnic and other important background factors of the examiner should be reported simultaneously with the test scores.

8. That the testing industry establish and fund an independent research and development corporation charged with the responsibility (1) to identify the critical problems in assessment as they relate to minority groups; (2) to sponsor research to investigate those problems requiring study; (3) to sponsor appropriate development work; and (4) to involve researchers who have the endorsement of minority professional and community associations, and that a minimum sum of four per cent (4%) of the income over expenses for non-profit testing corporations and the sum of four per cent (4%) of profits for profit-making testing corporations be set aside in support of the above objectives.

9. That test publishers exercise an advocacy responsibility which require that test objectives be stated clearly, that the process be fully described so that subjects and users understand exactly what is expected and how it will happen. This is the principle of informed use and informed consent.

10. That publishers of tests state, with clarity, on all descriptive information concerning a test, the specific uses for which the test is designed, the specific limitations of the instrument, and that they provide information as to how the results should be interpreted in acceptable professional practice.

11. That we call upon the appropriate professional associations, specifically the Association of Black Psychologists (ABPsi), American Psychological Association (APA), the American Education Research Association (AERA), the Association for Non-White Concerns of the American Personnel and Guidance Association (ANWC), and other appropriate groups to establish minimum standards for those who administer and interpret such standardized tests as tests of intelligence, aptitude, achievement and personality, and to develop standardized basic information about test administrators and the testing environment, to be provided routinely with any test scores.

12. That the NAACP mount a concerted effort to identify instances of testing abuse which call for legal remedies.

13. That the NAACP establish a national task force for the purpose of developing specific guidelines for laymen's participation in and support of any standardized assessment procedures, and that the NAACP insure dissemination of the guidelines to the broadest possible audience.

14. That citizens urge state elected officials to pass legislation to establish a task force for the development of an independent Office of Consumer Affairs for Testing and Student Evaluation.

15. That the Association of Black Psychologists design and conduct workshops around the Task Force Reports which will include, but not be limited to, the effects of race of examiner on test taker; differential validation and reliability; improper use of IQ as dependent variables in research projects, biases in test construction; the problem of misinterpretation and the development of alternative means of assessment.

ON THE USE AND MISUSE OF TESTS

Competent scholars have always recognized the possibility that intelligence tests and aptitude tests might be abused. Beginning with Binet, and including such later scholars as Ells, Klineberg, Allison Davis and such contemporary scholars as Robert Williams and Leon Kamin, they have called attention to potential and actual abuses. Nevertheless, the abuses have continued. Many black and other minority children have been unfairly stigmatized or inappropriately grouped for instructional purposes. Culturally sophisticated teachers and psychologists have protested that children who have been thus identified as "less able," "retarded," or what might be called "the six-hour mentally retarded," seem to lose that retardation immediately at the end of the school day. But back in the classroom the next morning, they conform to the predictions of those who first sequestered them in "special" classes.

The racist history of the testing movement is documented in Leon Kamin's The Politics and Science of I.Q., as well as in the writings of others. (See List of References appended). The current misassessment of blacks and others is rooted in a long--and often unscientific and malevolent--history. In this Task Force report, we are concerned with tests of "aptitude," "achievement," "personality,"--both individual and group--and any other method of assessment of individuals which results in ranking, sorting out and invidiously comparing.

It should be the function of assessment, including testing, to facilitate the development of human resources. To be acceptable in the field of education, testing programs should reveal student growth in skills, attitudes and understanding. To be used in connection with selection for employment, tests must measure variables which have something to do with the job for which an individual is being tested. These are the criteria against which all tests and testing programs must initially be measured: every professional testing procedure, method, technique, instrument or material--and the environment in which the tests are administered--must conform to these standards. That is the initial judgment by which any testing program must be evaluated, even before it is put into use.

It is important to note here that we are very much in favor of, and recognize the need for, appropriate and competent assessment. However, the existing standardized tests are unacceptable. They fall short of meeting three fundamental criteria (in addition to that of relevance, noted in the previous paragraph). Tests and testing procedures must predict accurately what they promise (predictive validity). Tests must measure adequately the content of the area they purport to cover (content validity). The testing program must be capable of leading to prescriptions which result in positive growth for the person being tested (prescriptive validity). It is abusive to misassess with an inadequate instrument. Equally important, it is abusive to continue, year after year, to use testing programs which have proved themselves to be tools which are either irrelevant to student progress or which actually, by predicting failure, induce malevolent results. Current aptitude testing or other testing processes and practices which result in the misassessment of blacks and other minorities result in educational mistreatment. All such tests must therefore be examined, not only for cultural or ethnic bias, but more importantly, in the light of their intended ultimate use. For all these reasons, we assert that fundamental questions must be asked about testing, beginning with the motivation of the movement and continuing through the assumptions which underlie the use of the instruments, raising questions about the accountability of the testing industry when measured against the educational outcomes for students.

In the light of the foregoing general considerations, the following recommendations are made:

RECOMMENDATIONS TO THE TESTING INDUSTRY

We recognize that a number of theoretical problems in assessment need attention. Among these are the issues of norm versus criterion-referenced testing; predictive versus prescriptive validity; aptitude versus achievement testing; content versus contextual validity; status versus process assessment; and the extent to which some less readily measured goals will be disregarded if standardized testing is given principal priority and support. (Knowledgeable members of the teaching profession will recognize that each item in this list calls for book-length treatment).

In addition, underlying all of these theoretical issues is a basic one, the question of "universality" in human behaviour. When a single standardized test is administered to an entire population, the underlying assumption must be that it is fair to all because all are "alike" in their possession of the qualities and information to be tested: they differ only in the degrees to which they possess these qualities or have the information. Such an assumption cannot legitimately be made for the entire population of the United States. In the light of the information that we have about differences between and within cultural groups within this nation, we believe the assumption of "universality" to be a serious error. Since there is no standard experience, the "standardized" test poses a serious problem of theoretical difficulty. When applied to the problem of testing minorities, this difficulty is greatly increased.

The foregoing concerns by no means exhaust the list; but this weekend conference can do no more than suggest the magnitude of the problems before us, without presuming to include everything. We know that existing tests have been built at costs that run into millions of dollars; yet, even with such an investment, we do not have tests that work as they are supposed to work. The truth is that the task of devising a full remedy is beyond the competence of a weekend conference, but the burden of proof for the utility and satisfactoriness of any testing procedure rests with the producers. We will have discharged our function in this Task Force if we point to inadequacies and indicate general directions of needed improvement.

The testing industry is (or ought to be held) responsible for the development of assessment procedures which conform to professional standards. These standards are provisionally described in Standards for Educational and Psychological Tests published by the American Psychological Association Inc. in 1974.

Within these general prescriptions, the minimum responsibility of the testing industry is to include within any standardized tests of ability, achievement, personality, or any other assessment procedure, specific consideration of and data related to three kinds of validity-predictive, content, and prescriptive (as we noted above). These data must be expressed as a coefficient or other appropriate systematic expression which is developed as a result of an adequate validation study. Such data must accompany all administrative manuals for use with standardized tests. Furthermore, in instances where standardized assessment results in a disproportionate sorting of groups according to ethnicity, the test developer must provide separate validity coefficients for the ethnic groups to which the assessment procedure is to be applied.

In addition to the foregoing, it is also the responsibility of the producers of assessment procedures to describe probable main effect-variables in the instructional setting, in standardized terms, which must be considered along with the results from standardized testing if the interpretations of test results are to be meaningful or acceptable. For example, we know that the ethnic background or level of skill of those who administer or interpret tests often has a major effect on the announced results of the program. Therefore, we have the right to expect that the testing industry will devise satisfactory and systematic ways of taking such effects into account.

Agencies make money from the administration of tests. Therefore, they have the responsibility to finance the measures necessary to the correction of their testing programs. Among these measures must be the inclusion, within the test-construction process itself, of persons drawn from culturally diverse backgrounds and of various ethnic identities. The evaluating experts must include representatives who are acceptable to the minority professional groups and other major community groups if the testing program is to escape its present image as being unduly weighted in favor of the dominant forces of American society.

The testing industry must establish and fund an independent research and development corporation charged with the responsibility (1) to identify the critical problems in assessment as they relate to minority groups; (2) to sponsor research which will investigate these problems; (3) to sponsor appropriate development work; and (4) to involve researchers who have the endorsement of minority-rooted professional and community associations.

In view of the enormous profits which have been made over the years by the testing industry, we recommend that some small redress of past errors be made by the voluntary application of four per cent (4%) of the net income of non-profit testing corporations and four per cent (4%) of the net profits of profit-making testing corporations, to the support of the foregoing objectives.

Our next point is that the makers of tests must be accountable for the uses to which their tests are put. Test publishers must be responsible for monitoring the use of their tests by assuming an advocacy role when necessary. In instances where tests are used (or are about to be used) without due observance of this advocacy role, test publishers should apply sanctions, including the denial of the use of their product by those who misuse it.

Moreover, test publishers have an advocacy responsibility which requires that test objectives be clearly stated, that the process of administering the test be fully described so that subjects and users both will understand exactly what is expected to happen and how it will happen, in non-threatening, affirmative terms. This we would call the principle of informed use and informed consent.

Finally, publishers of tests must state with clarity, in all descriptive information concerning a test they publish, the specific uses for which the test is designed, the specific limitations of the instrument, and full explanation as to how the results should be interpreted.

All of the foregoing steps must be taken by the testing industry itself. The broadest interpretation possible should be made of the concept of abuse, within the field of standardized testing, since abuse can occur at any and all points of the testing process, from initial development and conceptualization through utilization and interpretation.

We assert here that the misuse of tests, whether due to ignorance or to bias or to indifference, is an important factor in the total problem we are discussing; but we also assert that it is by no means the whole problem. Inherent in the instruments themselves, particularly in the instruments designed to measure aptitude, is a basic bias which must be corrected at its source. That source is within the offices and work-rooms of the testing industry. No amount of training or orientation of users of the tests will correct this built-in bias; but a successful effort to correct such bias will result in high predictive-, content-, and prescriptive-validity.

RECOMMENDATIONS TO PRACTITIONERS

Valid standardized tests, even when they are produced by the industry, can still be used in error, or their full potential lost. Appropriate use would include diagnostic and prescriptive procedures which lead to pupil gains. In the instance of educational tests designed to measure achievement, recent innovations in standardized assessment seem to offer possibilities which may be supportable.

Some recently devised tests would seem to have a greater utility for use in the school room than earlier tests. Those with the greatest utility have a closer relationship to the instructional process, offering valuable insights as to teaching strategies. These tests are often referred to as "formative-summative" tests. A common use of these tests occurs in reading and mathematics programs, where short "locater tests" are used to ascertain proper beginning points for individual students in appropriate units of instruction, and "summative" tests are used to ascertain the degree of mastery of materials and the readiness of the child to move on a more advanced level.

While no specific existing tests can be fully endorsed at this time, the principle involved in the assessment procedure just described appears to merit our support.

Practitioners are reminded that abuses in testing programs often occur with reference to majority as well as minority children. While we stress the pattern of abuse with reference to minority children, because the magnitude of the error is greater, we do not overlook the welfare and progress of the majority child--who should also benefit from a testing program designed to meet the criteria and serve the purposes we are discussing.

Any test which results in a disproportionate distribution of students on an ethnic basis must demonstrate that this disproportion does not come as a built-in error due to low predictive-, content-, or prescriptive-validity. Where such a conclusion cannot be established, persons asked to participate in the use of the test should decline to do so.

Perhaps an example will help to explain what we mean. Persons familiar with the processes of test construction will admit that items which, in the preliminary testing and validation, appear to differentiate between males and females are thrown out, presumably on the assumption that there are no real differences in intellectual functioning as between the two sexes. Yet items which appear to differentiate between whites and blacks, in the same battery of tests, are retained--presumably on the assumption that there are real differences between the races. Thus, the test which is published and administered, after the preliminary run, carries no built-in bias as to sex but does carry a built-in bias as to race. The fault is clear. The correction ought to be equally clear.

We call upon the appropriate professional associations, specifically the Association of Black Psychologists (ABPsi), the American Psychological Association (APA), the American Educational Research Association (AERA), the Association for Non-White Concerns (ANWC), of the American Personnel and Guidance Association, and other appropriate groups to establish minimum standards for those who administer and interpret standardized tests such as tests of intelligence, aptitude, achievement and personality. These standards must be equivalent in rigor to the standards which have been established for publishers of standardized tests and diagnostic techniques. Further, we call upon these same associations to develop standardized basic information about test administrators and the test environment, to be provided routinely with any test scores.

RECOMMENDATIONS TO THE NAACP

Recent court decisions (Larry P. et.al. v Wilson Riles, et.al.; Diana, et.al. v California State Board of Education) have been won on the principle that predictive validity for a standardized test was lacking in the instances invoked. In the case of Griggs v. Duke Power Co., the case was won on the principle that employment tests must be "job related". To maintain the momentum generated in these cases, it is recommended that other instances which may appropriately be pursued to a legal remedy be immediately identified and pursued.

Particular attention should be paid to the results flowing from the use of the National Teachers Examination. A disproportionately high number of those failing this examination come from the minorities. Teaching certificates are denied or credit for promotion is withheld as a result. Yet there is little, if any, evidence to indicate that these examinations have any predictive relationship to the jobs sought. The principle of Griggs v. Duke Power Co. would appear to apply.

If predictive validity cannot be demonstrated by the Law School Admissions Test (LSAT), the Graduate Record Examination (GRE), the Medical College Admissions Test (MCAT), the Graduate Management Admissions Test (GMAT), and other screening devices used in connection with admission to graduate and professional schools, technical schools or other institutions of higher education, legal redress should be sought.

We call upon the NAACP to establish a Task Force for the purpose of developing specific guidelines for the participation by laymen in the effort to devise satisfactory assessment procedures. Further, we call upon the NAACP to insure dissemination of acceptable guidelines to the broadest possible audience.

RECOMMENDATIONS TO LAYMEN

No person should consent to participate in an assessment program unless he or she has a reasonable understanding of the procedures to be applied, and of the outcomes to be expected.

RECOMMENDATIONS TO OFFICE HOLDERS

We urge that the public obligation to protect consumers against improper exploitation and victimization be recognized by legislative action to establish, outside the educational bureaucracy, a properly staffed Office for Consumer Affairs (Testing and Student Evaluation).

ON THE PSYCHOMETRIC INTEGRITY OF TESTS

INTRODUCTION

The purpose of the Task Force on Psychometric Integrity of Tests was to highlight a number of questions and issues central to the technical development of tests. There are many techniques and procedures in test construction that must be followed to ensure that an appropriate measure has been devised. Although most of us know very little of the technical skills required in building a house, we place faith in the constructor's ability to provide a safe and secure home. If too many flaws occur in its construction our lives may be placed in jeopardy and we must question the ability, perhaps the integrity, of the constructors. Over the years a plethora of regulations have helped govern the construction industry as a guaranteed protection of the citizen from fraud and misrepresentation of quality of product. Comparable to the building-trade industry is the test-construction industry, whose technical procedures must also be scrutinized for quality of product. Decisions made using test results oftimes place the lives and futures of children and adults in jeopardy. That is to say that test data must provide constructive information which will enhance the personal development of all individuals regardless of group membership. To date tests have principally been used as instruments for screening individuals and placing them in classes of opportunities. However, tests must also safeguard the individuals' potential and capacity to grow, if not specifically reveal ways in which they may be developed (e.g., in perhaps the diagnostic-prescriptive manner). This Task Force therefore raises questions about important technical areas of test construction which must be considered in the totality of issues on cultural fairness in testing for minorities. These questions represent areas for more complete analysis and discussion.

The Charge to the Task Force listed a number of questions which must be fully addressed by competent experts in the technical procedures of test construction and utilization. Our response cannot be definitive, given their complexity and our shortness of time. However, our intent is to share with others our awareness that part of the problem in fair testing of minorities rests with the technical development of tests as well as their use. Each issue, therefore, will be treated separately with a statement reflecting, in our judgment, an expansion of the issue in terms of minority concerns as well as suggesting the areas which need further inquiry.

1. As in most things, tests should have a reason and purpose underlying their development. Although attention may be given to the purpose for which tests are used, there is an equally important problem to consider--namely, the theoretical assumptions about human performance (usually intellectual performance) and about how the test format purports to measure and actually does measure the areas of interest. For example, one would assume that to test for intelligence a concept of what "intelligence" is and its behavioral manifestations would be clear. Technically, it is not. Intelligence has become synonymous with scores on tests of intelligence (i.e., the Intelligence Quotient or IQ). The full range of

intellectual potential and capacity of humans is still unknown. We know even less about how humans process information. Therefore, it is a gross injustice to discredit a person's present intellectual capacity, much less potential, solely on the basis of test scores which are in turn frequently based on a fragile foundation of theoretical understanding and empirical evidence about human intellectual ability. The tendency is to take what the majority of people can do as indicative of ability levels. However, the presupposition is that all persons have had at least equal access to the experience central to the skills of the majority. Nevertheless, within the issue of theoretical concepts explaining behavior, we must ask whether these basic assumptions consider the breadth of life contingencies (particularly minority human development within a racist society) which might affect intellectual growth. Certainly minority scholars have had little opportunity to participate in this domain of scholarly speculation, much less positive reception and support of their ideas and assumptions about human performance.

It is therefore necessary to scrutinize fully the assumption underlying many of these tests to see whether the root of unfairness in testing stems from the initial conceptualization of the behavior one is trying to measure, and ultimately how it is incorporated into a standard test format.

2. Many tests, particularly intelligence tests, have several parts (sub-tests) which, when taken together, are supposed to represent a total picture of a person's abilities. The same stringency of controls employed, in general, for the overall development of a test applies also to the establishment of criteria in the selection and use of parts of a test. For example, commonly used subtests in a test require demonstration of one's verbal, quantitative, or performance abilities. Certainly we agree that these components of ability are important in intellectual competence, but we are also concerned about the tendency for test developers and clients to neglect other domains of human capacity which affect performance (e.g., motivation, personality, emotions, memory strategies, etc.) Our concern, therefore, is to question the exclusivity of priority areas in ability testing and to emphasize the need to broaden the areas of human capacities considered.

3. Another area of important concern to us is the process of determining test questions which are ultimately selected to be a part of the instrument. It is acknowledged that this is an arduous process of sifting through numerous questions until a final group of items best represent the purpose of the test. Moreover, it is acknowledged that the statistical process involved has immense utility. However, the judgments in the development of questions and the interpretation of statistical data in the process of selecting items is still a human one, it can be no better than the competence and sensitivity of the experts assigned to this stage of test construction. Therefore it is important that at this juncture of test development we are clear about the criteria of the excellence we are trying to measure, and how widely that excellence is represented within the pluralistic cultural milieu of American society.

It is precisely the fact that American society is culturally pluralistic which concerns us, in the item selection process as well as other stages in test construction. Many items are selected on the basis of their alleged ability to measure the skills of the average person, the population in general. Very often this is at the expense of those items which differentially distinguish between various groups. In other words, a question which blacks may do well on, but not necessarily other groups represented in the tested population, may be excluded from the final form of a test because of its lack of representativeness within the general population. This may also be true of any other group distinctions one wants to make. The fact is that in the pursuit of questions which represent the general population, we may be overlooking information which has a comparable potential for representing human capacities, but in a selective manner. Moreover, excluding items from a test which favor one group but not necessarily another, may be placing in jeopardy the representation of that group's ability and potential.

4. The factor analytic procedure is a statistical method for finding out how a pool of items cluster together or are truly independent. It is a procedure central to the item selection process in test development. The arguments for and against this procedure and its dominance in determining test items is recognized as too technical to entertain within this document. The important point, however, is the recognition of the fact that this procedure has limitations; and, secondly, that there are few minority professionals involved with this procedure of test development as participants in judging the use of this statistical technique in the best interest of minority groups. There is a critical need to determine how much this statistical procedure may influence the type of test items selected, and what impact decisions by statisticians may have on the way minority ability is profiled on standard tests.

In addition, since there is considerable dependency on this procedure in test construction, we again question whether there are not other methods which might be comparable in objective, but more sensitive to the way different groups exhibit their abilities. In essence, we request a broadening of methods of test construction and assessment, particularly adopting those which would provide a fair representation of individual and group skills.

5. One area of testing which has been controversial is the issue of standardization procedures, the development of norms for a given population. Essentially, this involves profiling the range of competence demonstrated by a population on a set of questions comprising a test. These are ultimately refined into standards to which levels of individual achievement are to be compared. Central to this procedure is to profile the range of test performance from a representative assortment of individuals and groups. It has always been a concern of minorities that as a group we have not been adequately represented in the normative populations. Recently, some testing agencies have attempted to revise their norms by including a broader representation of minorities in the sample (e.g., Weschler, *Intelligence Scale for Children, Revised* (WISC-R).) There is still controversy as to whether proportionate representation will resolve the issue of fairness. The selection process should include those factors (e.g., socio-economic and geographical residency) that are

appropriate for the representative group which is to be tested. If the major concern is to have a test which is representative of the general population, the majority group in this country (whites) will prevail in the performance results of the test. Consequently, the peers to whom one's ability is compared are essentially representative of the standards set by individuals or groups outside the target populations.

6. Two major concerns about tests are whether they are reliable and whether they are valid. By reliability is meant the consistency of a person's performance on a test if repeated; that is, confidence that the level of an individual's or group's performance will not vary significantly the second time around. Validity of a test focuses on the truth or accuracy of test results, the degree of faith that can be placed in the test as measuring what it claims to measure. The importance of these dimensions for the issue of test fairness for minorities rests with the interdependency of these two factors. A test should be both highly valid and highly reliable. This is not always the case. A test with little validity may be reliable (i.e., one's performance may be consistently high or low in a test-retest circumstance), but the content or predictive capability may be unrelated to the test objective or the desired basis for judgments. Parenthetically, this is an issue in the IQ controversy.

Similarly, a test may wrongly be considered valid if it meets only one or some criteria--that is, if it has content validity, predictive validity, concurrent validity and/or construct validity. In the testing controversy, discussion tends to center on predictive and content validity. The issue of "culturally biased tests" principally refers to the fact that the contents of tests are not representative of the socio-educational experiences of minorities. The predictive validity of any test can really be no better than the defined future for which probable success is being predicted (i.e., success in school or on the job). The numerous issues within the meaning and significance of validity and reliability of tests (particularly as germane to the use and abuse of tests) requires that this area not be treated lightly. The public must fully understand the consequences in the assessment capabilities of a test when it does not meet acceptable standards of validity and reliability.

7. Cut-off scores and criterion variables are two other important factors in the testing process. By cut-off scores is meant the point at which a person's performance (a test score) permits a judgment about acceptance or rejection, that is, his or her ability or potential in terms of likelihood of success. Related to cut-off scores is the criterion variable or the performance goal. This refers to skills which a person or institution establishes as representative of acceptable performance. For example, a criterion variable for college admissions is "potential grade point average". This is purported to be indicative of college success. It may be determined that scores at a particular level on a test correspond to a projected level of performance in college (i.e., expected GPA). Consequently, as in the case of admissions tests, e.g., College Boards (SAT) or Graduate Record Examination (GRE), a person's score supposedly serves as a predictor of the level of achievement he or she will attain. By looking at the relationship between the admissions tests scores of students who

are successful in college, schools establish their "cut-off scores" as well as criterion variables.

The Task Force's difficulty with this process is that prediction models of potential success build a dependency on two factors -- the test score and the institution's definition of "success" -- with little consideration for other factors which might contribute to a person's success (e.g., motivation). As previously noted, if there are serious questions about the validity of a test, then the significance of the test scores is brought into question. Furthermore, if test scores are used in a process of selection based on prediction, the criteria of "success" depends upon the institution's valid appraisal and inclusion of all salient factors that represent competent performance. Very often institutions provide a narrow definition of success or skills required (e.g., success in college equals potential grade point average).

Many minorities are denied access to educational opportunity because of institutional dependence on poorly determined levels of cut-off scores, or are enrolled in institutions where, if traditional criteria of selection had been employed, they would have been excluded from this opportunity (a case in point is the City University of New York's Open Admissions Policy).

8. Because standardized tests are reasonably short, economical and easy to administer, they have become the most convenient method of assessment. The Task Force's concern is that this dependency on such a structured format has limited the exploration of other means of assessing intellectual ability. We are concerned not only with the limited domain of demonstrated competence on which tests tend to focus, but also the limited strategies of assessment (e.g., standardized pencil and paper tests) employed by institutions. It seems unfair to judge the obviously broad capacities of human growth and development in such a narrow manner. Likewise, given the breadth and variety of human social experience it seems unfair to build dependency on measures which actually test only for degrees of conformity (in knowledge and experience) among the general population -- not innate abilities. The judgments about one's ability and, subsequently, the consequence for one's life are too important to be limited to performance on tests. Other alternative means for measuring and predicting human potential and capacity must be developed and employed. The multitude of assessment strategies must reflect the pluralistic compositions of American society. This is no easy accomplishment, but at the same time this objective must not be ignored at the expense of someone's life opportunities.

9. The importance of how a test is used is an enormous issue in and of itself. Within this Report, another Task Force is discussing this problem alone. Although the content of that Task Force's Report covers this issue in depth, it cannot be overemphasized that the qualifications of the examiner or test user are equal to, if not more important than, the tests which are used, since even the best instrument in the hands of unqualified users can lead to disastrous results. No mechanism exists at present to insure that only well-qualified persons (or agencies) will use the tests or establish policies and procedures based upon test results.

10. The Task Force on Use and Misuse of Tests speaks specifically to the current objective of "criterion-referenced" tests. Therefore, this Task Force will address itself to a related assessment methodology, i.e., development of tests which are a derivative of the social and cultural experiences of the specific group to be tested. Tests of this nature are considered "culture-specific tests". The utility of these tests has been hotly debated; but in the Task Force's judgment they serve a purpose -- if only in delineating the context and content of learning experiences for that particular group. In this regard, it is believed that efforts to develop "culture-specific tests" as an inductive procedure ultimately advances our knowledge from specific bases of information to those common features which describe the general population. With the current test-industry effort to find the domains of convergence in excellence and performance of a general population (which in effect standardized tests do), little weight is given to the idiosyncratic nature of learning. It is this Task Force's belief that tests standardized both for the general population and for specific populations add to the information on variations in the acquisition of knowledge. Within a pluralistic society, we must know what people have in common, but also where they differ, without discrediting either in the quest for understanding. To seek a test which is "culture-free" in content is nice in theory, but to date has proven impractical. There is very little that can be identified which is not influenced by the cultural context in which it is nurtured and expressed. In general, the Task Force endorses the efforts to develop and refine criterion-referenced and culture-specific tests as part of alternative assessment strategies.

In conclusion, the Task Force on the Psychometric Integrity of Tests recognizes that there are many technical issues involved in the construction of tests. These must be considered within the context of their impact on fairness in testing minorities. Because of the complexity of many of the areas discussed within this report, it should not be construed that all points have been raised -- much less exhausted. The Task Force has tried to highlight what constitutes some of the major concerns presented to minorities by the technical process in test construction. As in so many other professional areas, minority professionals are badly under-represented in the testing industry. In particular, the contribution of minorities within the institutions and agencies which develop tests and set policy has been negligible. The test development process has remained untouched by an external system of public accountability. The consequences to one's life from performance on tests can be as pivotal as any lifesaving drug. In the latter instance, there is public accountability (i.e., FDA), in the former, there is none. The following recommendations of this Task Force are a step in the direction of public accountability of the testing industry.

RECOMMENDATIONS

The Task Force asserts that the basic goals and purposes of assessment are as appropriate for minority individuals as for persons in the majority group. It is the consensus of the Task Force that there can be a meaningful and worthwhile place for the testing function in the assessment of individuals. However, the issues which concern the NAACP, and in fact, the purposes of this Conference, relate to the constraints and restrictions placed on minorities by testing; by the fact that those negative results preclude access to educational and occupational advancement; by the absence of social and cultural considerations in test construction; and by individual and group values which affect test performance and interpretation.

The Task Force's interpretation of the Resolution adopted by the 65th NAACP Annual Convention, is that it serves as a means for identifying the need to investigate the concerns of minorities in testing. It has also provided opportunity for study which may devise better ways of developing and using test data.

SPECIFIC RECOMMENDATIONS

1. Normative procedures and specification must be carefully developed to ensure fairness to the test-taking population. Information of this type should not be used for minority groups or individuals without appropriate norming (study) on that population.

2. The test development process must consider the different cognitive structures and styles of different groups. Studies related to minority test performance indicate that group differences do affect performance. Factor analytical methods may yield information relevant to understanding these cognitive structures and styles, but there still remains the problem of how these factors are related to effective performance.

3. The test selection-predictive system should include other variables, e.g., motivation, persistence, "creativity", and other personality measures. There is frequent mention of these important factors in the discussion of minority assessment, with little follow-through.

4. The relationship between the time factor and test results should be expanded with minority groups. This exploration should not be limited to test speediness, but also to the length of time covered by the criterion measure.

5. The test administration process, both for individual and for group testing situations needs to be monitored to ensure quality control of tests results. We recommend that guidelines be developed that will provide opportunities for optimum test performance.

6. Culture-specific tests should be considered an integral part of the test construction process.

7. Noting the difficulty of establishing the relevant criteria, particularly where predictive validity strategies are used, it is strongly recommended that culturally-appropriate and content-valid, criterion-

referenced procedures be investigated and adopted. Some guidance should be provided in the development of criterion variables, e.g., tests for firemen, policemen, etc.

8. The development of alternative strategies in assessment is strongly encouraged.

9. There should be a thorough review of current tests being used. NAACP, ABPsi, and others should actively pursue the possibility of seeking funding sources for such a project.

10. There should be a concerted effort by testing industries and professional schools to recruit and train more minority persons in psychometric techniques. Furthermore, testing industries should increase participation of minority professionals in the test development process.

ON PUBLIC POLICY

INTRODUCTION

We live in a highly diverse, highly competitive, credential-oriented society where success and winning have become so prized that often they become ends within themselves. As a result of the frenetic milieu created by this kind of attitude and philosophy on the part of most Americans, an increasing number of assessment tools and techniques are being employed to facilitate the classification, stratification and certification of individuals and groups in our society.

The fact that testing and the results of testing (assessment) have had (and continue to have) a strong impact upon public policy is not new. A review of litigation involving the alleged misuse of tests by certain businesses and certain industries and the results of judgments in those cases indicated clearly the role that testing has played in hiring practices and the impact that they were seen to have on policy. The cases to which we allude here involved white plaintiffs.

The determination of public policy related to the testing of minorities (and more particularly American blacks) has been insidious and extremely deleterious. The history of our country is replete with evidence of so-called scientific material designed to show the inferiority of the Negro. The Police Reference Notebook states, "A large body of literature came into existence to prove that the Negro was imperfectly developed in mind and body, that he belonged to a lower order of man, that slavery was right on ethnic, economic and social grounds..." The rationale for counting slaves as three-fifths of a person for determining the number of representatives that a state might send to Congress was based on "scientific" tests that "proved" the inferiority of the Negro. For centuries blacks were disfranchised by the use of "tests." It is unnecessary to belabor these historic truths; it is important to realize, however, that testing helped weave the racist fabric of the United States. There are daily, painful reminders that white racism (the unfair treatment of nonwhite persons, based solely on skin color) continues to flourish in the United States. The insidious character of white racism (and the most dangerous component of its insidious nature) is most harmful when it becomes an inextricable part of test construction. This situation can obtain without the conscious participation of the developer or publisher. Nevertheless the disastrous results, however unintentional, are assured. When public policy is based on an already abusive instrument and the interpretation of an insensitive researcher, persons belonging to minority groups are exposed to a variety of inequitable and unethical behaviors and treatments.

THE PROBLEM

The Task Force on Public Policy, in its concern about the impact of tests on public policy, felt it incumbent upon its members to consider various facets of the problem. That is to say, the person who uses the test and the individual who formulates public policy are as important as are the tests, in the overall consideration. Naturally, the publisher must assume his part of the total responsibility for fair and equitable public policy.

Some of the psychometric instruments having the greatest impact upon public policy have been:

1. Tests of "intelligence" (I.Q. Tests)
2. Personality Tests
3. Placement Tests
4. Achievement Tests
5. American College Testing Program (ACT)
6. Scholastic Aptitude Test (SAT)
7. Miller's Analogies
8. Graduate Record Examination (GRE)
9. Interest Inventories
10. Graduate Management Admissions Test (GMAT)
11. Law School Admissions Test (LSAT)
12. Medical College Admissions Test (MCAT)
13. Other Admission or Aptitude Tests

Certain questionnaires have also been responsible for questionable public policy, but this Task Force could not consider them as a part of this report.

RECOMMENDATIONS

After careful study of the charge to the members of the Task Force, and following in-depth deliberations, the Task Force on Public Policy makes the following recommendations:

1. There should be a moratorium on all current standardized tests, unless these instruments conform to recommendations submitted by the other Task Forces that comprise the Conference on Minority Testing.

This recommendation is based on studies which tend to prove that most standardized tests are inherently racist (intentionally and unintentionally) and that they do discriminate against minorities. The precedent for calling for this moratorium has been set by (a) the National Association for the Advancement of Colored People (NAACP), (b) the National Education Association (NEA), (c) the Association of Black Psychologists (ABPsi), and (d) the National Association of Elementary School Principals (NAESP).

This call for a moratorium does not mean that all forms of assessment should be abolished. Some form of assessment is and will always be necessary. It does recognize the damage that has been done and is being done as a result of public policy based on the spurious results obtained by the use of questionable instruments, as well as by men whose motives and preparation for their work are questionable.

The over-representation of minorities in Special Education classes across the country and the many cases of litigation resulting from the misuse or abuse of I.Q. tests gives further credence to this recommendation. The labels resulting from the misuse of I.Q. tests and the damage to the victims of these labels are evident in our public schools.

Loss of financial assistance both for programs and for individuals have resulted from public policy based on student achievement. A case in point is the Head Start Program that actually was very successful, but was judged on ill-conceived measures and improper assessment tools. Other programs have suffered budget cuts or have been eliminated on the basis of test results, when it was the test that was the failure, not the program. The same may be said for many students who have been denied financial assistance for their education based largely or solely on test results.

The Graduate Record Examination has a very poor predictive record but universities continue to use it as an exclusionary instrument. Standardized Admissions tests for entrance to college and professional schools have such questionable predictive validity for certain segments of our student population that a moratorium would seem to benefit all applicants -- not just black applicants.

There are sufficient studies to support this position. Some of these can be found in the appended bibliography. It is both unethical and inappropriate to base public policy on any results obtained through the use of current I.Q. tests on minority groups.

2. It is recommended that a national monitoring body, with the power to enforce, through sanctions, be established to assure proper assessment and policy regarding the administration of assessment tools.

Without such a monitoring body, there is no guarantee that even legislation related to ethical testing procedures will be honored. This group should be a national body and should, by composition and sensitivity, reflect the best interests of our various minority populations along with those of the majority. Selection procedures may be determined after further deliberation.

A test or other assessment tool is no better than its user. We are concerned about what may be called the "experimenter variable," i.e., even the value of an acceptable instrument may be destroyed by one who is not qualified professionally or personally. The individual who assumes the responsibility of interpreting the test results of minority group individuals must be sensitive to the nuances of the many cultures within our pluralistic society. An insensitive person can contaminate test results or adversely affect the individual being tested by even unconscious manifestations of his insensitivity. This position, too, is based on research in this area.

3. Companies that develop, publish, and sell tests must assume (or continue to assume) a major responsibility for assuring the corrections of ills related to their product. This Task Force has strong feelings about this issue, and concurs with the other Task Forces on the mandatory nature of this obligation.

4. The appointment to public office or the nomination and election to office (particularly as that office involves the establishment of public policy based upon the testing process) should always be based upon an individual's knowledge of, sensitivity to, and ability to converse with all segments of society, since all will be affected.

5. No individual currently in public office should make decisions about public policy based on the results of testing or research on minority groups without the concurrence of groups, organizations or individuals most knowledgeable of and conversant with life-styles, value-attitudes and "experience-in-America" of those minority groups.

ON A FAIR TESTING CODE

PREAMBLE

In a purported meritocratic and adversarial milieu in which so much credence is placed upon education, and in all of the values and status derived from it, the need for evaluative criteria and assessment tools is prominent. From the cradle to the grave there has been an increasing reliance in this country on the use of an assortment of standardized tests to screen, select, admit, reject; to classify, stratify, track, license or certify. Such subjective dependence on "objective" tools makes it possible, and even attractive in some quarters, to use them inequitably in ways which preclude a segment of the population from reaching the heights of its aspirations and realizing its potential, or from acquiring the skills and financial resources to develop their own.

In order to prevent the further misuse of tests or other measuring devices, to dispel existing myths about their infallibility, to ensure a fair and equitable use across all segments of the population and to promote the intended best use of standardized tests, the testing industry and those who subscribe to their services must become more introspective about the impact of the testing phenomenon.

Further, the testing industry must assume a greater responsibility in correcting the abuses and misuses that result from application of their products (tests), which attempt to measure intelligence, ability, aptitude, achievement, and other potentials, matters which are critically important for participating in the mainstream of society. The Task Force on the Code for Tests and Testing calls upon both the industry and the user, not only to engage in introspection, but to adopt the code which follows, and to be governed by it as they continue to work with tests and test-related activities.

CONSTRUCTION OF TESTS

Tests, regardless of their claims of objectivity, are a reflection of the experiences, characteristics, values and training of those individuals who construct them. Many of these tests are then used in the decision-making process which determines the future of minority people* who are excluded from the developmental process. Those tests which are administered widely to minority and majority people alike, and which attempt to measure intelligence, achievement, aptitude and to predict the potential of all test-takers, must employ trained persons who have experienced life as a member of a minority. Minority representation must be involved in the process of conceptualizing and developing tests. More specifically, minority persons must be involved in the overall development from its initial conceptualization to the final product. As an alternative, minorities may be forced to find financial resources and to develop their own testing programs.

---

* Minority people as used in this Code designates the following: Black, Spanish speaking (Puerto Ricans, Hispanics, Chicanos, Latinos); Native Americans (Indians); Asian-Americans (Japanese, Chinese, other orientals).

STANDARDIZATION PROCESS

Since most test scores are interpretable only in relationship to the group on which the test was normed or standardized, and since the norming process is so critical to the entire concept of standardized tests, the testing industry is called upon to define and make public in prominent, clear and appropriate literature the process by which its tests were standardized. Further, since there exists an assumption that blacks and other minorities are usually at or below the norm of their white counterparts, it is imperative that the norms reflect the pluralistic characteristics of the different ethnic groups that make up the tested population. Not only should minority people be included in the norming population, but the testing industry must identify the sample characteristics on and by which the test was standardized.

CONDITIONS OF ADMINISTRATION

Inasmuch as the test results of individuals frequently have immeasurable influence on their status in life, the testing industry and those who administer tests should ensure that optimal and uniform conditions always prevail. For example, where centers are established for local administration of nationwide tests consideration should be given to logistical problems, e.g., the distance candidates are required to travel. Buildings should be properly placarded so that students who are unfamiliar with the location of testing rooms, rest rooms, and other critical areas are not disadvantaged. Testing rooms should be well lighted and ventilated and should contain writing surfaces which are comfortable for all. Proctors should be hired to reflect the ethnic make up of the candidates being tested. All Proctors must be sensitive to the needs, questions and/or anxieties of all candidates.

No individual may be permitted to administer tests to members of minority groups or to interpret such data unless (a) he is duly qualified and proficient in the technical aspects of the testing process and (b) he can demonstrate a keen sensitivity to the life-styles, value-attitudes and "experience-in-America" of the several populations being submitted to the assessment procedures.

In summary, there should be a regulatory mechanism which not only monitors test administration, but one which takes the necessary corrective actions to proscribe irregular and unfair administration of tests.

LIMITATIONS OF USES

Failure to adhere to a code of conduct or to regulate oneself, on the part of the test developer and the test user could not only escalate the now rampant misconceptions that exist about standardized tests, but also invite external regulation of the testing industry. The testing industry, accompanied by the test user, must take the lead in divesting itself of the misconceptions about what tests can and cannot do, and then in dispelling similar misunderstandings among the public-at-large, and ultimately in promoting a more diagnostic, cautious and creative use of tests in the educational process, to wit:

a. Intelligence Tests (or ability tests) -- there is a notion held by many test users that intelligence or aptitude is synonymous with an immutable or fixed characteristic within an individual. This fixation, or one's "native ability," is said to determine what is expected of one, and also one's level of expectation for all time. The testing industry must describe and publicize the fact that its intelligence or aptitude tests do not measure, in an interpretable manner, one's level of expectation throughout life. The industry must be more forthright in calling attention to the fact that learning depends not only on inherited abilities but importantly, also upon life experience in a particular environment. Since what is learned may differ according to one's economic status in life, the tests should not be used as a predeterminer of the level to which an individual aspires and may obtain.

b. Admissions tests -- myriad tests are produced and used in the admissions process to post-secondary and to graduate institutions. Because of the nature of the admissions process, many of these tests are used not to admit but to exclude. The testing industry must define the proper use of admissions tests, and explicitly state the conditions under which tests should not be used. The extent to which a test is intended to predict one's performance at a given level dictates the extent to which the instrument must be validated on the entire entering population. The testing industry is obliged to encourage strongly such validation of admissions tests and to perfect a model or mechanism for effecting it.

c. Occupational and Professional Tests -- tests for occupational and professional certification or entry should be job-related. A major issue in the test-selection process is whether the test measures abilities appropriate to performance in the job sought. The lack of correspondence between test requirements and job requirements invalidates the test (e.g., using broad aptitude or achievement tests for hiring firemen or policemen). Fitting the test to performance in the job has become the current objective of "criterion-referenced tests". In a job selection situation this approach makes sense. Its utilization in other assessment contexts needs further exploration.

The U.S. Supreme Court in Griggs v. Duke Power Co. held, "If an employment practice which operates to exclude Negroes cannot be shown to be related to job performance, the practice is prohibited." Griggs v. Duke Power Co. and Chance v. New York have indicated that there must be a match (bonafide relatedness) between the test content and the skills and knowledge needed in performing a job. Where there is a mismatch between test content and expected performance, a serious barrier is established for the person who is subjected to the tests.

The Task Force on the Code for Testing calls on the testing industry to refrain from establishing contractual relationships which require variance from guidelines suggested by the courts.

CLEAR INTERPRETATION

Since misconceptions of tests can lead to damaging uses, the testing industry is obligated to make full disclosure of the purposes for which tests are designed, the processes by which they were designed, the population on which they were standardized, the statistical characteristics which delimit their use, e.g., standard error of measurement, standard error of estimate, and other "do's and don'ts" which will affect optimal use and interpretation. More specifically it is incumbent upon the testing industry to give clear and precise interpretation of the scores their tests yield. The public must be informed not only that the test scores are fallible and that their reliability is imperfect, but must be told the extent of that fallability and imperfection. The public must be informed that test scores are only a sample of a student's performance and are never more than an estimate of truth.

CONSTRUCTIVE USES OF TESTS

Effective uses can be made of tests, regardless to their type, only if the user knows what the test contains, what its purposes are, and what its limitations are. We move on the assumption that an educational system, and particularly a school within the system implicitly guarantees that students to whom achievement examinations (tests) are administered have been taught in ways such that they can reasonably be expected to have learned the information required by the examination. Therefore, we call on the testing industry to assist school systems and other users in understanding better the content and constraints of the examinations, and in helping them understand how to make optimal use of the results. Moreover, we call upon the educational community and the public and private sectors to ensure that tests are not used or relied upon where such guarantees do not exist.

RESEARCH

We recognize that many people question whether any tests have credibility. With regard to the validity of some tests the evidence is inconclusive, even among the strongest advocates of testing. Recognizing that some device is going to be used to determine access to institutions, jobs, professions, and other opportunities within society, and to determine the various ways in which human resources are used, we assert that any testing program which results in a significantly disproportionate distribution of scores by ethnicity must meet the most stringent validity requirements. Therefore, continued test use must be contingent upon intensified and continued research on the effects (both negative and positive) of tests, on the educational opportunities and related problems of blacks and other minorities.

Some of this research needs to be reactive in the sense that it addresses things that have occurred, and some should look ahead to what ought to be done to circumvent certain problems. Quality of the research, credibility of the researcher and methodology continue to be essential elements. The need exists for more than the traditional methodology which has characterized the testing industry's efforts in the past, and, which, for an increased number of Human Resources agencies, has also become standard practice. Just as important as these components is the way in which data from that research are interpreted. We call on the testing industry, educational systems, and public and private sectors who have a vested interest in education, to hire and use black and other minority researchers who can assist in collecting, analyzing and interpreting that critical mass of data which can help us understand the effects of tests on the problems of minorities.

BETTER TRAINING OF USERS IN INTERPRETATION

Teachers, counselors and admissions officers are publics important to the testing industry since they sit at the entrance gates through which many test takers must go if they are to realize their goals. The extent to which these publics can correctly interpret and effectively use those instruments is the extent to which decisions will be made that are fair to minority test takers. The testing industry is called upon to help bridge that gap which exists between effective use and lack of understanding, by conducting workshops and institutes, and writing special publications that are aimed at interpretation of tests.

ACTION AGENDA

The recommendations for an Action Agenda are ordered according to the publics toward whom our recommendations are directed, i.e., the testing industry, the Association of Black Psychologists, the NAACP and laymen.

THE TESTING INDUSTRY

1. Develop and publish standards of competence for those who administer, score and/or interpret tests.

2. It is known that the technical information about the test varies from one cultural group to another, often in highly significant proportion. It is essential that the testing industry provide technical information appropriate for ethnic groups for whom the testing is done.

3. Establish and fund an independent research and development corporation to identify the critical problems in assessment as they relate to minority groups; sponsor investigative research and development work involving researchers who have the endorsement of minority group professional and community associations.

4. State with clarity on all descriptive information concerning tests they publish, the specific uses for which the test is designed, the specific limitations of the instrument and a full explanation as to how the results should be interpreted.

5. Establish a national monitoring body, with the power to enforce, through sanctions, to assure proper assessment and policy of assessment tools.

ASSOCIATION OF BLACK PSYCHOLOGISTS

1. Design a project to identify some of the major tests now operating to screen individuals out of educational and employment opportunities and develop a position statement on those tests, applying the Standards for Educational and Psychological Tests, developed by a Joint Committee of the American Psychological Association, Inc., American Educational Research Association and the National Council on Measurement in Education. The results of the Project to be published and disseminated widely - to users and to clients, e.g., major school districts, counselors, and the black helping professions such as social workers and nurses.

2. Develop an empirical demonstration project for refuting the kind of conclusions drawn from I.Q. and achievement tests.

3. Design and conduct workshops around the Task Force reports including effects of race of examiner, validation and reliability, improper use of I.Q. as dependent variables in research projects, biases in test construction, problem of misinterpretation and the development of alternative means of assessment.

NAACP

1. Develop a statement regarding the rights of clients, including whether an individual

has to take an I.Q. test; where they can go to get information regarding the validity of the test.

2. Initiate legal action to develop a body of case law on the testing issue.

3. Approach the American Psychological Association regarding the inclusion in the accreditation criteria, the evaluation of the capability of the psychology training program to produce special competence in cross-cultural assessment; standards for approving programs of training and curriculum offerings in the field of psychology.

4. Establish a National Task Force to develop specific guidelines for the participation by laymen in the effort to devise satisfactory assessment procedures and disseminate the guidelines on the broadest possible basis.

5. Develop a fact sheet or "Know Your Rights" pamphlet for parents and students regarding standardized testing.

6. Urge legislation at the state and federal level establishing, outside the educational bureaucracy, a properly staffed Office for Consumer Affairs for Testing and Student Evaluation.

7. Bring together a coalition of organizations to implement the recommendations of the Task Forces.

LAYMEN

1. Know your rights as a consumer. Urge the NAACP, the Association of Black Psychologists to speak to community groups regarding the rights of individuals participating in testing programs.

2. Urge your legislators at the State and National level to sponsor legislation to establish, outside the educational bureaucracy, a properly staffed office of Consumer Affairs for Testing and Student Evaluation.

SELECTED REFERENCES

Advisory Committee on Evaluation, in Goal, a publication of the Joint Committee on Educational Goals and Evaluations, Assembly, P.O. Box 83, State Capitol, Sacramento, Calif. 95814, November 2, 1974.

Ainsworth, M.D., Sensory motor development of Ganda infants. In *The Competent Infant*, New York: Basic Books, (1973), 503-14.

Allen, V.F., What does a Reading Test, Test? Philadelphia: Temple University College of Education. "TTT" Project Monograph Series, No. 1, USOE Contract No. OEG-0-70-2046 (721), (1974).

American Psychological Association, American Educational Research Association, and National Council on Measurement in Education (Joint Committee). "Technical Recommendations for Psychological Test and Diagnostic Techniques." Psychological Bulletin, 51, no. 2, Part II (1964).

American Psychological Association, American Educational Research Association, and National Council on Measurement in Education (Joint Committee). *Standards for Educational and Psychological Tests.* (1974).

Anastasi, Anne, *Psychological Testing.* 3rd ed. New York: The MacMillan Co. (1968).

__________Psychological tests: uses and abuses. *Teachers College Record*, 62, (1961), 389-393.

Assessing minority group children, *Journal of School Psychology* (Special Issue), Vol II, No. 4, (1973).

Baratz, S.S., and Baratz, J.C., Negro ghetto children in urban education. In *Linguistic-Cultural Differences in American Education*. The Florida Foreign Language Reporter, 1, (1969).

Bay Area Association of Black Psychologists. Position statement on use of IQ and ability tests. In *Black Psychology*, ed. R. L. Jones, New York: Harper & Row, (1972), 92-4.

Berger, P.L. and Luckmann, P., *Social Construction of Reality: A Treatise in Sociology of Knowledge,* Garden City, N.Y.: Doubleday and Co. (1966).

Black Advisory Task Force to the Child Development Associates Consortium. *Collaborative Assessment: A Position.* Washington, D.C.: The Child Development Associate Consortium (1974).

Black Intelligence Test of Cultural Homogeniety (BITCH), Author, Robert L. Williams, Washington University, St. Louis, Missouri (1972).

Brenner, A. and Stott, L.H., *School Readiness Factor Analyzed*, Detroit, Mich.: The Merrill-Palmer Institute (1968).

Brown, R., How shall a thing be called. In *Children with Learning Problems*, edited by S.G. Sapir and A.C. Nitzburg, New York: Brunner/Mazel (1973), 378-387.

Campbell, J.T., Tests are valid for minority groups. *Public Personnel Management*, 2, (1973), 70-73.

Clark, K.B. and Plotkin, L., *The Negro Student at Integrated Colleges.* New York: National Scholarship Service and Fund for Negro Students (1964).

__________A review of the issues and literature of cultural deprivation theory. In *The Educationally Deprived* ed. K.B. Clark, et al. New York: Metropolitan Applied Research Center (1972), 74-92.

Cleary, Anne T., Test bias: prediction of grades of Negro and white students in integrated colleges. *Journal of Educational Measurement,* 5, (Summer 1968), 115-124.

Cohen, R., The influence of conceptual rule-sets on measures of learning ability. *Race and Intelligence*. Washington, D.C.: American Anthropological Association (1971), 41-57.

Cole, M., Gay, J., Glick, J.A. and Sharp, D.W., *The Cultural Context of Learning and Thinking: An Exploration in Experimental Anthropology.* New York: Basic Books (1971).

Cole, N.S., Bias in selection. *Research Report No. 51*. Iowa City, Ia: American College Testing Program (1972).

Dove Counterbalanced General Intelligence Test. Author, Adrian Dove (ca. 1966).

Deutsch, M., Katz, I., & Jensen, A.R., Social Class, Race, and Psychological Development. New York: Holt, Rinehart & Winston (1968).

Farr, R. and Griffin, M., Measurement gaps in teacher education. Journal of Research and Development in Education, 7 (1973), 19-28.

Flaugher, R. L., Some points of confusion in discussing the testing of black students. ETS Research Memo, RM 73-75, Princeton, N.J. (March, 1973).

__________ The new definition of test fairness and selection; development and implications, rm Educational Testing Service, RM 7317, Princeton, N.J. (September, 1973).

Franklin, A.J., The testing dilemma for minorities, Teachers College, Columbia University. (ERIC Document Reproduction Service No. ED 103557) (1974).

Garcia, John, IQ: The conspiracy. Psychology Today, (Sept., 1972), 40-43, 92-97.

Gardner, John W., Excellence, Harper & Row: New York (1971).

Ginsberg, H., The Myth of the Deprived Child: Poor Children's Intellect and Education, Englewood Cliffs, N.J.: Prentice-Hall (1972).

Glasser, B.G. and Strauss, A.L., The Discovery of Grounded Theory: Strategies for Qualitative Research. Chicago: Aldine (1967).

Goodwin, Leonard & Joseph Tu., The social psychological basis for public acceptance of the social security system: The role for social research in public policy formation, American Psychologist, Vol 30, No. 9, (Sept., 1975), 875-883.

Griaule, M., Conversations with Ogotemmli: An Introduction to Dogon Religious Ideas, New York International African Institute -- Oxford University Press (1965).

Harris, T. George, An introduction: I.Q. abuse, Psychology Today, (Sept. 1972), 39.

Hilliard, A.G., The strengths and weaknesses of cognitive tests for young children, in One Child Indivisible, ed. J.D. Andrews, Washington, D.C., The National Association for the Education of Young Children (1975).

Hobbs, N., ed. Issues in The Classification of Children: A Source Book on Categories, Labels and Their Consequences. San Francisco: Jossey Bass (1974).

Hofstadter, Richard, Anti-Intellectualism in American Life, Alfred A. Knopf: New York (1974).

Jones, R., Black Psychology, New York: Harper and Row (1972).

Jones, R.L., ed. Labelling children culturally deprived and culturally disadvantaged. In Black Psychology, edited by R.L. Jones. New York: Harper and Row (1972), 285-294.

Jorgensen, C., Racism in mental testing: The use of I.Q. tests to mis-label black children. In Explorations in Psychology edited by A.A. Harrison, Monterey, Calif.; Brooks/Cole (1974).

Kamin, L.J., The Science and Politics of I.Q. Potomac, Md.: Lawrence Erlbaum Associates (1974).

Kirkpatrick, J.J., Ewen, R.B., Barrett, R.S., and Katzell, R.A. Testing and Fair Employment: Fairness and Validity of Personnel Tests for Different Ethnic Groups. New York: New York University Press (1968).

Klienfeld, J.S., Intellectual Strengths in Culturally Different Groups: An Eskimo Illustration. Review of Educational Research, 43, (1973), 341-60.

Labov, W., The logic of non-standard English. In Language and Poverty, Chicago: Markham, (1970), 153-189.

Lesser, G.S., Gordon, F., and Clark, D.G., Mental abilities of children from different social class and cultural groups. Monographs of the Society for Research in Child Development, 30 (4, Serial No. 102), (1965).

Levi-Strauss, C. The Savage Mind. Chicago: University of Chicago Press (1966).

Levenson, E.A., The Fallacy of Understanding, New York: Basic Books (1972).

Lewis, D. Anthropology and colonialism. Current Anthropology 14, (1973).

Linn, R.L., Fair test use in selection. Review of Educational Research, 43, (1973), 139-161.

Mercer, Jane R., I.Q.: The lethal label, Psychology Today, (Sept., 1972), 44-47, 95-97.

__________Sociocultural factors in the education evaluation of Black and Chicano children; Paper presented at the 10th Annual Conference on Civil and Human Rights of Educators and Students, Washington, D.C.: (Feb. 18-20, 1972).

Milholland, J.E., Culture-Fair Intelligence Test. In Sixth Mental Measurement Yearbook, Tests and Reviews: Intelligence Group, ed. O. Buros. Highland Park, N.J.: Gryphon Press (1965).

Moore, T.E., Cognitive Development and the Acquisition of Language. New York: Academic Press (1973).

Myrdal, G., Objectivity and Social Research. New York: Pantheon Books, (1967).

Nobles, W., Psychological research and the black self-concept; A critical review, J. Social Issues 29, 1, (1973), 11-31.

Rippey, R.M., ed. Studies and Transactional Evaluations. Berkeley, California: McCutchan (1973).

Roberts, S.O. and Shane, M.D., Selecting superior students for Fisk. Fisk News, 36 (No. 3 Spring), (1962), 5-6.

__________and Birch, J.S., College careers of superior Negro American high school graduates. Report to Carnegie Corporation of New York, from Fisk University (Department of Psychology) (1966).

__________and Oppenheim, D.B., The Effect of Special Instruction upon Test Performance of High School Students in Tennessee. Princeton, N.J.: Educational Testing Service (1966).

__________and Horton, C.P. & Roberts, E.T., SAT vs GRE performance of Negro American college students. Proceedings, 77th Annual Convention, APA, (1969), 177-178.

__________and Horton, C.P., The Relationship of Selected Non-Intellectual Factors to Freshman Academic Achievement. Unpublished manuscript, Fisk University (Department of Psychology), (1970).

Ryan, W., Blaming the Victim, New York: Vantage Press (1971).

Samuda, R.J., Psychological Testing of American Minorities. New York: Dodd, Mead & Co. (1975).

Senna, C., The Fallacy of I.Q. New York: The Third Press (1973).

Stanley, J.C., Predicting college success of the educationally disadvantaged. Science, 171, (1971), 640-647.

Stone, Charles C., Testing and the educational power struggle, Integrated Education: Race and Schools, Vol IX, No. 4, (July-August, 1971), 4-10.

Terman, L.M. and Merrill, M.A., Stanford-Binet Intelligence Scale: Manual, Etc. Boston: Houghton Mifflin (1960).

The Journal of Afro-American Issues, Vol III, No. 1, Winter, 1975, (Issue on Testing, Measurements and Afro-Americans).

The 1975 Statement of the American Psychological Association on Testing and Minorities and replies to the same.

Cleary, T.A., Humphreys, L.G., Kendrick, S.A. and Wesman, H.A. Educational uses of tests with disadvantaged students; American Psychologist, 30, (1975), 15-41.

Jackson, G.D., Comment on the Report of the Ad hoc Committee on Educational uses of tests with disadvantaged students: Another psychological view from the Association of Black Psychologists. American Psychologist, 30, (1975), 88-93.

Bernal, E.M., Comment - On the Report of the Ad hoc Committee on Educational uses of tests with disadvantaged students: A response to "Educational Uses of Tests with Disadvantaged Subjects": American Psychologist, 30, (1975), 95-96.

Humphreys, L.G., Addendum - To the Report of the Ad hoc Committee on Educational uses of tests with disadvantaged students: American Psychologist, 30, (1975), 95-96.

Cronbach, L.J., Five decades of public controversy over mental testing. American Psychologist, 30, (1975), 1-14.

Thomas, A. and S. Sillen, Racism and Psychiatry, New York: Brunner/Mazel (1972).

Thorndike, R.L., Concepts of culture-fairness. Journal of Educational Measurement, 8 (1971), 63-70.

Torrance, E.P., Gifted Children in the Classroom, New York: The Macmillan Company (1965).

Turner, L.D., Africanisms in the Gullah Dialect, New York: Arno Press, New York Times (1969), originally published 1949, University of Chicago.

Watson, Peter, I.Q.: The Racial Gap, Psychology Today, (Sept., 1972), 48-50, 97-99.

Webb, E.J., Campbell, D.T., Schwartz, R.D., and Sechrest, L., Unobtrusive Measures: Non-Reactive Research in the Social Sciences. Chicago: Rand McNally and Company (1966).

Weber, D.A., Racial bias and the LSAT: A new approach to the defluse of present admissions, Buffalo Law Review (1975), 439-462.

Wechsler, D., Manual for the Wechsler Adult Intelligence Scale, New York: Psychological Corporation, (1955).

Williams, E.B., Test Bias and Predictive Validity. Paper presented at the Annual Meeting of the American Psychological Association, Chicago (1975).

Williams, R.L., Abuses and misuses in testing black children. In Black Psychology, edited by R.L. Jones, New York: Harper and Row (1972), 77-91.

__________BITCH gives Blacks a break on tests. Norfolk Journal and Guide. Saturday, August 23, 1973, 11.

__________Black pride, academic relevance, and individual achievement. In Crucial Issues in Testing, edited by R.W. Tyler and R. M. Wolf. Berkeley, Calif.: McCutchan (1974), 13-20.

Test Bulletins

College Entrance Examination Board (CEEB), ATP-Student Bulletin, 1975-76. Princeton, N.J., 1975-76.

Educational Testing Service (ETS), Graduate Record Examination (GRE) Information Bulletin 1975-76. Princeton, N.J., 1975-76.

American College Testing Program (ACT), Medical College Admission Test (MCAT) Bulletin, Iowa City, Iowa, 1975-76.

Educational Testing Service - Law School Admissions (LSAT) Bulletin. Princeton, N.J., 1975-76.

Psychological Corporation. Bulletin of Information - List of test centers - (MAT) Miller Analogies Test and other tests, New York, 1974.

American College Testing (ACT), Seminars for ACT users. Iowa City, Iowa: ACT, Author, 1975.

Educational Testing Service (ETS), Programs of Continuing Education. Princeton, N.J.: ETS, Author, 1975.

Mr. WEISS. The next witness is Dr. Lionel Newsom.

STATEMENT OF LIONEL NEWSOM, MEMBER, BOARD OF DIRECTORS, NATIONAL ASSOCIATION FOR EQUAL OPPORTUNITY IN HIGHER EDUCATION, AND PRESIDENT, CENTRAL STATE UNIVERSITY (OHIO)

Dr. NEWSOM. Mr. Chairman, members of the subcommittee: My name is Lionel Newsom. I am president of Central State University in Ohio. I am also a member of the board of directors of the National Association for Equal Opportunity in Higher Education, whose acronym is NAFEO.

With me is our executive director, the former president of Bowie State, Dr. Samuel Meyers.

NAFEO is the membership organization consisting primarily of the historically black colleges. Our membership currently consists of 107 historically and predominantly black colleges that enroll more than 200,000 students and continue to graduate even today approximately one-half the black recipients of the baccalaureate degree.

I welcome this opportunity to speak to you on behalf of the NAFEO membership and the constituency it serves. A number of the presidents and chancellors of our institutions, through professional training and experience, qualify as experts on testing.

In addition, as educators, we have strong views on the wide spectrum of issues with which H.R. 3564, Truth in Testing Act of 1979, and H.R. 4949, Educational Testing Act of 1979, are concerned.

However, given the fact that extensive and, I would add, high quality testimony has been prepared by such groups as the United States Student Association and the American Council on Education on issues such as Government regulations, individual student rights, disclosure, reasonable costs, and the like, and given the fact that relatively little attention has been given to the possible impact of the proposed legislation on minorities and the economically deprived, we have chosen in our testimony to correct this imbalance by focusing attention on the implications of testing legislation for these groups.

After careful analysis and extensive discussion, it is our judgment that the enactment of H.R. 3564 would have a deleterious impact on minorities and should not be enacted and that H.R. 4949 should be delayed in passage, modified in some respects, augmented and expanded in others, and in yet others, changed in focus following a year-long study by a legislatively mandated commission on testing.

NAFEO concurs with and applauds those aspects of H.R. 3564 that recognize the importance of testing of scholastic aptitudes and achievements as a factor in college and university admissions, and of the objective of the proposed legislation to enhance equal opportunity and to insure equal rights and fair treatment of individuals applying for admission.

We also approve the recognition of the importance of testing of skills as a factor affecting occupational entry and the necessity for taking some action to assure that there is a free flow of skills in interstate commerce. We interpret this to be consistent with the goals of expanding the flow of all persons, including minorities, into occupations.

Our concern is that the legislative language will give implicit endorsement to testing as the principal factor in admissions and to testing as a critical factor in employment.

There are many of us who believe that multiple criteria, of which testing is only one, should be employed in determining college and university admissions and employment entry.

We endorse the proposal set forth in section 6 of H.R. 3564 to prohibit the conducting of educational and occupational admission tests unless such tests are administered in a manner to protect the rights of the individuals tested.

We also concur, in general, with provisions of section 6(a) that pretest information should be provided to applicants for educational or occupational admission tests. We also endorse, with one major exception, section 6(b) which requires that individuals who have taken an educational or occupational admission test should be notified of their individual specific performance.

We object to section 6(b)(3) requiring notification of the score required to pass the test for admission to such occupation or the score which is generally required for admission to institutions of higher education. Such a requirement would exaggerate the importance of tests which should be only one of a number of multiple criteria for admissions.

In addition, such a requirement would, through legislation, institute a practice of specifying a cutoff score for admissions, which does not now exist and would, thereby, erect additional college admissions and occupational entry barriers to minorities and the economically disadvantaged.

Many of the laudable objectives set forth in H.R. 3564 could be accomplished by building upon the current practices of testing agencies reinforced by some of the self-regulatory proposals set forth by others. The rather superficial, and often ignored, pretests distributed by testing agencies could be expanded, made more understandable, and built into workshops, institutes and curricul of institutions in order more thoroughly to promote an understanding on the part of applicants for college admissions. Programs could be implemented to develop similar information for those seeking occupational entry. The emphasis would, accordingly, properly shift somewhat to the use of testing as a diagnostic and developmental instrument.

In summary, we concur in the noble purposes of H.R. 3564 to promote truth in testing and to enhance educational opportunity and occupational mobility through improving pretest information on college admissions and employment entry tests and on reporting test results to the tested. We believe that these laudable objectives can be accomplished by strengthening and expanding upon existing mechanisms and through enhanced self-regulatory measures.

A greater concern is that if enacted in its present form, H.R. 3564 could exacerbate rather than resolve the problems of minorities and other low-income groups by exaggerating the importance of testing and by raising barriers to the admissions and entry of minorities and the economically disadvantaged to colleges and trades.

We are in similar concurrence with the laudable objectives of H.R. 4949, the Educational Testing Act of 1979. We endorse the concept that education is fundamental to the development of individual citizens and the progress of the Nation as a whole. This is particularly true for minorities and the economically disadvantaged.

We also concur that there is a continuous need to insure equal access for all Americans to educational opportunities of high quality. We also recognize that standardized tests are a major factor in the admission and placement of students in postsecondary education and also plays an important role in individuals' professional lives.

We concur with the other premises of this proposed legislation: Of recognition of the increasing concern about testing; protection of the rights of individuals; serving the public interest; and that standardized tests have a repercussion nationally.

We also concur in the purpose of the act: to improve the awareness of test subjects; of the characteristics, uses and limitation of admission tests; to make available to the public appropriate information about tests and testing; to protect the public interest by developing knowledge from research on tests and testing; and to encourage the use of multiple admissions criteria.

The thrust of the proposed legislation is that:

One. The test agencies would be required to provide a great deal of specific information to the test subjects before and immediately after the taking of the tests.

Two. The test agencies would be required to report to the Commissioner of Education statistical data and reports pertaining to the tests that would form the basis of an evaluation required of the Commissioner of the impact of the tests.

Three. There would be full disclosure of test results and of the test itself, although privacy rights would be protected. There would also be detailed reporting required on testing costs and fees to students.

Four. Finally, implementation of the act is called for within a 120-day period after the effective date of the act with promulgation of regulations authority accorded the Commissioner of Education. Penalties for violators are also specified.

Our interpretation is that primary emphasis among those reacting to the proposed legislation has focused on the alleged abuses of the testing agencies. The agencies have been characterized as monopolist-like giants victimizing helpless students, whose entire educational careers or life's professions have been jeopardized by a force testing, over which they have had no control.

We recognize also that test agencies contend that the disclosure of test questions will invalidate the tests and, thereby, greatly increase test preparation costs and, thus, costs to students. We also recognize that others refute the contention that test preparation costs will necessarily rise.

We are convinced that costs would increase and, thereby, have a disproportionately adverse effect on low-income students. We are even more greatly concerned by a corollary effect that we consider likely. We believe that the dissemination of the test questions will spawn the development of more coaching schools. Minority students, who are disproportionately overrepresented in low-income groups, would generally be unable to attend these coaching schools.

We believe that coaching can improve test performance to some degree. The net impact of the increase in the cost of testing and the overrepresentation of the affluent in coaching resulting from the disclosure provisions of the act would be adverse to the interest of minority and poor students.

The proposed legislation gives inadequate attention to the misuse and abuse of test. Our impression is that sophisticated professors, admissions officers, and others who are in a position to observe both test scores and the actual performance of the test subjects, soon lose their enchantment with the inviolability of tests. Experts in the testing agencies appear keenly aware of the limitation of the tests they administer.

On the other hand, lay groups, possibly including prospective employers, tend to have an exaggerated impression of the importance of testing. The proposed legislation directs its attention, primarily, to the supply side, control of the test agencies. It is possible that more consideration should be given to the use of cautions, that are now so familiar on hazardous substances, to warn the users of the dangers and limitations of tests.

More corrective action directed toward the demand side would seem appropriate, but is not considered in the current legislation.

Even though misused and abused, testing has unfulfilled potential. At the present time, standardized tests together with prior academic grades, family characteristics, recommendations, and similar factors predict with fair accuracy, the academic performance for 1 year in the future. They cannot predict long-term academic performance or long-term success in professional careers. Much more research is needed to improve the performance of the test itself as a long-run productive instrument.

Perhaps the most important area overlooked by the proposed legislation concerns the possibility of improving the performance of blacks and other minorities on standardized items.

Researchers at a session on testing at a conference recently held by NAFEO stressed that all black students do not have low scores on standardized tests and all educationally disadvantaged students are not black. Nevertheless, it was pointed out that black students entering the Big Ten institutions tend to score on the ACT from 1 to 1½ standard deviations below the mean of the entering freshman class for the institution.

We know of some historically black colleges that take students whose test performance is poor and develop them to the level that they are above the average at some of the Nation's most prestigious graduate schools, and they prove to be successful, productive professionals in our society.

In spite of these success stories, there seems to be a movement to expand the use of test to exclude low-scoring subjects from higher education channels without concomitant effort being exerted to improve the performance of these students on tests.

It should be stressed that a systematic and cumulative exclusion of any group from educational channels that lead to the benefits of a productive society will deny that society the use of potentially productive resources, but will concomitantly create, at some point, a mass of malcontents who will threaten the tranquillity and strength of that society.

It is accordingly vitally important to use the research and the success stories that we have to identify coaching techniques, curricula changes, life experiences, and overall strategies to improve the performance of blacks and other minorities on standardized tests. Legislation merely designed to regulate the testers will not accomplish this broader goal.

Our major recommendation, accordingly, is to delay enactment of H.R. 4949. We strongly urge that a commission, established by legislation, should be formed so that within a 1-year period it would carefully consider proposed legislation that would change the emphasis from regulating testers and, thereby, increasing costs, and adversely affecting minorities to a more positive emphasis.

The new proposed legislation would: One, carefully consider the research that could enable us to get even greater potential from standardized tests; two, outline a whole series of hazardous labels to caution users against the abuse of tests and; three, chart strategies to improve performance of minorities and the poor on such tests.

We would hope that the Commission were accorded subpena powers to assure availability, at least for confidential use, of research data already available.

Mr. Chairman and members of the subcommittee, we commend to you our recommendation to delay passage of H.R. 4949 and to appoint a commission to provide factual data so that the proposed legislation might be modified, augmented, expanded, and refocused along the lines we have proposed.

In doing so, your subcommittee, we predict, will fully utilize this opportunity to resolve some fundamental issues concerning testing that affect the strength and viability of our Nation.

Mr. WEISS. Thank you, Dr. Newsom.

Next we will hear Mr. Sheldon R. Goldstine.

STATEMENT OF SHELDON R. GOLDSTINE, CLEVELAND, OHIO

Mr. GOLDSTINE. Thank you, Mr. Chairman.

I would appreciate the opportunity to express my gratitude for being invited to testify at this hearing today, September 24, 1979, concerning H.R. 3564, the Truth and Testing Act of 1979, and H.R. 4949, the Educational Testing Act of 1979.

My testimony is in favor of both bills oriented toward H.R. 3564 based on my architectural testing experiences, legal involvements, and research into the architectural licensing process.

My academic and professional background is as follows: Academic: B.S. Architecture, the Ohio State University, 1972.

B.S. Business Administration, the Ohio State University, 1972.

Master of Business Administration, Case Western Reserve University, 1974.

Master of Architecture, the Ohio State University, 1975.

The combined business/architectural academic program was designed and implemented to supplement my architectural education (typically, most architectural schools throughout the country are extremely weak in education of basic business principles) to enable me to efficiently and effectively cope with the economic demands being placed on the building industry today. Per one of my applicant references (exhibit C-2), (all exhibits referred to in oral testimony retained in subcommittee files) submitted to the State Board of Examiners of Architects in Ohio on October 11, 1976, by a former Ohio State University architectural graduate school professor and former president of the Architects Society of Ohio (State component of the American Institute of Architects), I am regarded as one of the top students overall to graduate from Ohio State University, and highly qualified to be brilliant in the profession.

Professional: I have been involved professionally in the architectural field since 1973 and have experienced all conventional phases of the practice.

I and thousands of other applicant architects throughout the country have suffered severe professional hardships trying to gain entry into the architectural profession by means of the current National Council of Architectural Registration Boards (NCARB) professional exam.

Personally, I have tried, three times unsuccessfully, to pass the NCARB examination. An exam written entirely by NCARB's Professional Examination Committee aided by the Educational Testing Service (ETS) of Princeton, N.J., in question organization and format.

The NCARB, Washington, D.C., is a nonprofit, membership organization established under the laws of Iowa, comprising the architectural licensing boards of all 50 States, the District of Columbia, the Canal Zone, Guam, Puerto Rico, and the Virgin Islands.

It was established in 1920 to aid in the interstate practice of architecture. A certificate issued by NCARB to an architect entitles him/her to be admitted to practice in each of the 55-member board jurisdictions.

The 100-percent multiple choice NCARB professional exam, first given in 1973, is offered only once, in December of each year. Ohio and several other States have required a 1-day graphic/written exam in addition to the NCARB professional exam.

The multiple-choice exam is currently administered by all 55 NCARB member boards, machine graded by ETS, Berkeley, Calif., results sent to NCARB, Washington, D.C., for a national pass-fail cutting-score determination, and then forwarded back to all 55 member boards for their distribution of results.

Several States have modified the NCARB cutoff scores slightly; however, all candidates must meet the NCARB cutoff score guidelines to be eligible for NCARB certification, which is necessary to compete in the more profitable national architectural building market. Also note, several States are currently taking definite steps to eliminate and replace the current NCARB professional exam.

Per a phone conversation with Mr. Hal Levin, a member of the California Board of Architectural Examiners, on September 17, 1979, his board has been funded by the California Legislature to eliminate and replace the current NCARB exam by December 1980. California will offer their new exam twice per year.

Per another phone conversation with Mr. Cass Hurc, former executive secretary of the Wisconsin Architects-Engineer Board, also on September 17, the Wisconsin board is definitely taking steps to move away from the NCARB exam with pressure from the Wisconsin Legislature derived from a series of exam analyses dating back to 1973. The Wisconsin Legislature grew concerned over the abnormally high failure rate among NCARB examinees.

The most recent Wisconsin study, based on a task analysis by Wisconsin practicing architects, to my understanding, proved the NCARB exam invalid. For years, States throughout the country requested, without response, that NCARB effectively analyze its exam. The NCARB exam will be discussed in more detail later in this presentation.

In regard to my testing and administrative appeal experiences, I will present a case documenting the lack of basic constitutional requirements for due process of law.

Shortly after receiving a failure score on the December 1976 NCARB exam, I appealed in a meeting before the Ohio Board, so that they could add a learning element to the testing process.

The board's coercive authoritarian relationship maintained between itself and applicant architects is not conducive to meaningful cooperative negotiations to learn from past examination mistakes. Test questions and answers are not available for review and standardized reporting of scaled percentage score results do not allow for effective performance analysis.

That particular meeting between myself and the board in early 1977 was to me intimidating and probably did more harm to myself than good since I questioned the status quo of exam and significant level of performance feedback secrecy.

Again, in December of 1977, I failed the NCARB exam. Still, unaware of various State attacks against the validity of the exam, I felt the exam per se had many problems. Therefore, since I believed the exam was inadequate, was unable to get relevant performance feedback and felt I could make a significant contribution to the advancement of the profession, I applied for licensure without examination under section 4703.08(A) Ohio Revised.

A hearing was held before the State board of examiners on May 18, 1978. The board and myself were represented by legal counsel. Please note, prior to the hearing, my Ohio graphic/written design problems, two failures and one passing were destroyed in violation of section 4743.02 Ohio Revised Code.

The board's executive secretary stated in the hearing transcript that he was not aware of that provision. I am not calling this destruction of evidence an intentional act.

However, I would like to note that the design portion of the multiple choice NCARB professional exam is the only section that I have not passed, by 2 percentage points.

Also, per the trade magazine, Building Design and Construction, "A/E Licensing Under Attack in Several States," May 1979, p. 15,". . . when a legislative committee had the (Wisconsin) board ask a panel of Wisconsin architects to regrade portions of the December 1973 NCARB exam, the results of that regrading correlated badly with the official results, which led the State to eliminate the (graphic) design portion from the State's architectural exam."

Evidence was presented by myself at the May 18, 1978, hearing documenting my 6-year master of architecture degree and satisfactory completion of 3 years of experience. In regard to the satisfactory evidence of knowledge of professional practice and supervision of construction which the board at their discretion could also require, I called the board's executive secretary 1 week prior to the hearing in front of my attorney's law clerk to get directions.

The executive secretary stated that he did not know what would be required per that phone conversation and this also was documented in the hearing transcript.

The board's executive secretary, in response to their legal counsel's questioning, testified at the May 18, 1979, hearing, that to his knowledge no one had ever been admitted to practice under section 4703.08(A) O.R.C., and a list of approximately 350 approved applications was documented (see exhibits C-12 through C-16).

Unfortunately, the above facts could not enter the court as evidence since the board's denial of certification was upheld by the court of common pleas and was already in the court of appeals process.

The analysis of senate bill No. 15, as passed by the Senate, Ohio's 113th General Assembly, regular session 1979–80, a bill which added further restrictions to architectural professional entry, paraphrases section 4703.08(A) O.R.C. and states that it is one of the two basic methods of establishing proficiency as specified in the law to obtain architectural certification.

However, also documented in this analysis is the following statement:

"According to the board this method was used to 'grandfather' into the system architects practicing at the time the law passed, but is not currently used by the board to grant certification", by reviewing the history of section 4703.08(A), it becomes clear that the actual 'grandfather' clause, (sec. 1334-7(C)), was repealed in 1932 and that section 4703.08(A) was still a viable alternative according to the Ohio Legislature at the time I applied.

The board denied my application under section 4703.08(A) O.R.C. with the justification "that the applicant did not provide any evidence at the hearing which would show that he possessed any special qualifications for licensure without taking the examination required under section 4703.07, Ohio Revised Code" and "that the applicant at the time of the hearing had not passed the examination required by section 4703.07, Ohio Revised Code.

In rebuttal, the statute only calls for satisfactory evidence and the facts presented strongly indicated that the board had determined prior to the hearing that no matter what evidence I submitted, their willingness to license under the "license without examination statute", a separate alternative statute for licensure, would be negative.

The presented facts included the analysis of Ohio Senate bill No. 15 indicate, in my opinion, that an abuse of discretion did occur. It is most difficult to spend over 1 year in administrative appeal proceedings with the conceptual framework of the matter predetermined before the presentation of facts eliminating probability of true due process of law.

One of my major justifications for presenting my personal case analysis here is to illustrate the need for this Congress to design a standardized due process conceptual framework of professional testing guidelines to insure that all matters of business handled by State licensing board provide for a continuous process of due process.

In regard to the critiques of NCARB professional exam per se, I will briefly present information from relevant research data accumulated.

"An evaluation of the examinations administered by the architects section of the Architects, Professional Engineers, Designers and Land Surveyors Examining Board," submitted to the State of Wisconsin, Department of Regulation and Licensing by Dr. Glenn E. Tagatz of International Personal Services, Inc., Milwaukee, Wis., in February 1976, made the following findings:

> The NCARB examinations do not meet the specifications of the American Psychological Association with regard to use of psychological measurement technology in exam preparation with respect to job skills; not the Equal Employment Opportunity Commission's guidelines for job relatedness; nor the Uniform Guidelines on Employee Selection Procedures of the Equal Employment Opportunity Coordinating Council.

The "content analysis of the NCARB examination", exhibit B-2, prepared for the State of Wisconsin, Department of Regulation and Licensing, prepared by Dr. Peter E. Schriber, Mr. Michael Priestly, and Mr. Brent Campbell of National Evaluation Systems, Amherst, Mass., in June 1979, concluded the following:

> The most important conclusion is that the match between NCARB examination content and the Wisconsin Task List was less than 50 percent on an item-by-item match basis. Also, the proportionality of the NCARB examinations differed considerably from that of the Task List.

Please note that "EEOC guidelines state that test content should be as close to 100 percent job-related as practical and that the further away from 100 percent the less job-related the test is."

I do not want to imply that NCARB's use of its current exam is an intentional act to restrain trade. However, an only 50 percent job-related exam may in effect do so. The exam is graded on the basis of relative distribution of scores. The national pass rate is approximately 54 percent more or less each year.

Per the 1978 NCARB Architectural Registration Handbook, published by the National Council of Architectural Registration Boards and Architectural Record Books, 1978, page 14:

> All of the score distributions for the Professional examination have relatively small standard deviations, that is, the dispersion of scores is only about one-half of what might be expected in tests of this length. This would suggest that the examinees who actually took the test tended to be relatively homogenous, i.e., quite similar in terms of both personal characteristics and ability, an assumption which is borne out by the description of candidates in terms of age, educational level, et cetera.

Since the exam is less than 50 percent related and score distributions have relatively small standard deviations, guessing well on exam questions seems to be the primary vehicle into the profession.

Another problem must be presented concerning the lack of uniformity in applicant "intern" education. The intern period ranges from 1 to 5 years depending on college degree status, 13 years in some States for nondegree candidates, with 3 years being the average for 5-year professional architectural degree holders, bachelor of architecture.

The problem concerns the diversity of architectural practices in both their size and the building-type markets they are involved in.

Many intern architects are only exposed to work areas that return maximum profits to their employers.

Without a uniform intern training program, incentives will continue to exist which motivate employers to exploit this captive "cheaper-labor intern" labor force rather than providing the kinds of training these employees are paying for in reduced wages and, in effect, increase their chances, assuming a valid exam is given, of passing their architectural licensing exam.

Mr. Patrick Clinton's article: "A/E Licensing Under Attack in Several States", published in the May 1979 issue of Building Design and Construction, pages 14-15, 16, provides an excellent summary of current NCARB validity problems across the country.

Proform's article, "Wisconsin Architects Balk at National Exam, Design Their Own," published by the National Center for the Study of Professions, Washington, D.C., November 1978, gives an excellent account of activities in Wisconsin leading up to NCARB examination elimination and replacement.

Mr. Ron Shattil's article: "A Critique of the Registration Examination as a Measurement of Beliefs Rather than Ability," American Institute of Architects Journal, November 1976, and letters to NCARB from concerned parties give good accounts of content problems associated with the NCARB exam.

Illustrations of recent concerns expressed by Federal agencies regarding the negative economic effects of our increasingly expanding professional cartels are found in the following statements:

"An Address by Michael Pertschuck, Chairman, Federal Trade Commission, before the American Enterprise Institute Licensure Conference," Washington, D.C., February 22, 1979; and "Occupational Licensing: Consumer Protection or Cartel Management?" remarks by Mr. Donald L. Flexner, Deputy Attorney General, Antitrust Division, Department of Justice, before the Kentucky Legislative Research Commission's Program on Occupational Licensing, Lexington, Ky., February 2, 1979.

The statements made by Mr. Pertschuk and Mr. Flexner illustrate our need to be concerned about professional cartels limiting competition to keep prices high. Mr. Pertschuk stated one of the problems is the State licensing of professionals, a process usually dominated by members of the professions who hold down the number of entrants to control competition.

I have been cooperating with the Federal Trade Commission in investigation of the NCARB exam since March 1979. Also, I have been following, with much interest, efforts by professional political action committees aimed at inhibiting the FTC's regulatory authority which is directed toward maintaining a competitive economic system that does not unreasonably constrain private economic activities.

I am hopeful that this Congress will carefully analyze the tradeoffs of their actions toward the FTC with respect to the balance of our mixed system of government and private enterprise as well as our mixed system of monopoly and competition.

A major issue to be evaluated is the direction of our administrative legal system. Unlike our civil and criminal systems, which have developed with us over the centuries, the administrative system was born early in this century. Unchecked, it could have strong negative effects on our economic and political system.

I pray that our American economic dream of growth and property continues for centuries into the future. Insuring that professional entry is maintained at reasonably competitive levels in consideration of providing for the public health, safety, and welfare is but one component of the chain of events that will insure the American dream.

Again, thank you for this opportunity to testify before you today.

[The complete statement of Sheldon Goldstine follows:]

Sheldon R. Goldstine
2570 Overlook Road
Cleveland, Ohio 44106
(216) 371-0408

September 24, 1979

Honorable Carl D. Perkins, Chairman
and Membe.
Committee on Education and Labor
Subcommittee on Elementary, Secondary,
and Vocational Education
CONGRESS OF THE UNITED STATES
HOUSE OF REPRESENTATIVES
B-346C Rayburn House Office Building
Washington, D.C. 20515

Dear Chairman and Members of the Subcommittee on Elementary, Secondary, and Vocational Education:

I would appreciate this opportunity to express my gratitude for being invited to testify at this hearing today, September 24, 1979, concerning H.R. 3564, the Truth and Testing Act of 1979, and H.R. 4949, the Educational Testing Act of 1979.

My testimony is in favor of both bills oriented towards H.R. 3564 based on my architectural testing experiences, legal involvements and research into the architectural licensing process.

My academic and professional background is as follows:

ACADEMIC

B.S. Architecture, The Ohio State University, 1972.

B.S. Business Administration, The Ohio State University, 1972.

Master of Business Administration, Case Western Reserve University, 1974.

Master of Architecture, The Ohio State University, 1975.

The combined business/architectural academic program was designed and implemented to supplement my architectural education, (typically, most architectural schools throughout the country are extremely weak in education of basic business principles) to enable me to efficiently and effectively cope with the economic demands being placed on the building industry today. Per one of my applicant references (Exhibit C-2), submitted to the State Board of Examiners of Architects in Ohio on October 11, 1976, by a former Ohio State University Architectural graduate school professor and former president of the Architects Society of Ohio (state component of the American Institute of Architects), I am "regarded as one of the top student overall to graduate from Ohio State University" and "highly qualified to be brilliant in the profession".

PROFESSIONAL

I have been involved professionally in the architectural field since 1973 and have experienced all conventional phases of the practice.

I and thousands of other applicant architects throughout the country have suffered servere professional hardships trying to gain entry into the architectural profession by means of the current National Council of Architectural Registration Boards (NCARB) professional exam. Personally, I have tried three times unsuccessfully to pass the NCARB examination. An exam written entirely by NCARB's Professional Examination Committee aided by the Educational Testing Service (ETS) of Princeton, New Jersey, in question organization and format.

The NCARB, Washington, D.C., is a non-profit, membership organization established under the laws of Iowa, comprising the architectural licensing boards of all fifty states, the District of Columbia, the Canal Zone, Guam, Puerto Rico and the Virgin Islands. It was established in 1920 to aid in the interstate practice of architecture. A certificate issued by NCARB to an architect entitles him/her to be admitted to practice in each of the fifty-five member board jurisdictions.

The 100% multiple choice NCARB professional exam, first given in 1973, is offered only once in December of each year. Ohio and several other states have required a one day graphic/written exam in addition to the NCARB professional exam. The multiple choice exam is currently administered by all fifty-five NCARB member boards, machine graded by ETS, Berkeley, California, results sent to NCARB, Washington, D.C. for a national pass-fail "cutting-score" determination, and then forwarded back to all fifty-five member boards for their distribution of results. Several states have modified the NCARB "cut-off" scores slightly, however, all candidates must meet the NCARB "cut-off" score guidelines to be eligible for NCARB certification, which is necessary to compete in the more profitable national architectural building market. Also note, several states are currently taking definite steps to eliminate and replace the current NCARB professional exam.

Per a phone conversation with Mr. Hal Levin, member of the California Board of Architectural Examiners, on September 17, 1979, his board has been funded by the California legislature to eliminate and replace the current NCARB exam by December, 1980. California will offer their new exam twice per year.

Per another phone conversation with Mr. Cass Hurc, former Executive Secretary of the Wisconsin Architects-Engineer Board, also on September 17, 1979, the Wisconsin Board is definitely taking steps to move away from the NCARB exam with pressure from the Wisconsin Legislature derived from a series of exam analyses dating back to 1973. The Wisconsin legislature grew concerned over the abnormally high failure rate among NCARB examinees. The most recent Wisconsin study, based on a task analysis by Wisconsin practicing architects to my understanding, proved the NCARB exam invalid. For years, states throughout the country requested, without response, that NCARB effectively analyze its exam. The NCARB exam will be discussed in more detail later in this presentation.

In regard to my testing and administrative appeal experiences, I will present a case documenting the lack of basic constitutional requirements for due process of law.

Shortly after receiving a failing score on the December, 1976 NCARB exam, I appealed in a meeting before the Ohio Board so that they could add a learning element to the testing process. The Board's coercive authoritarian relationship maintained between itself and applicant architects is not condusive to meaningful cooperative negotiations to learn from past examination mistakes. Test questions and answers are not available for review and standardized reporting of scaled percentage score results do not allow for effective performance analysis. That particular meeting between myself and the board in early 1977 was to me intimidating and probably did more harm to myself than good since I questioned the status quo of exam and significant level of performance feedback secrecy.

Again, in December of 1977, I failed the NCARB exam. Still, unaware of various state attacks against the validity of the NCARB exam, I felt the exam per se had many problems. Therefore, since I believed the exam was inadequate, was unable to get relevent performance feedback and felt I could make a significant contribution to the advancement of the profession, I applied for licensure without examination under Section 4703.08 (A) Ohio Revised Code (Section 1334-7 (A) General Code). The statute read as follows:

> Sec. 4703.08. The state board of examiners of architects may, in lieu of all examinations, accept satisfactory evidence of any one of the qualifications set forth under the divisions of this section:
>
> (A) A diploma of graduation from an accredited architectural school or college showing that the applicant has completed a technical and professional course of not less that five years' duration, which course is approved by the board, and, in addition thereto, has had at least three years of satisfactory experience two years of which shall have been in the office of a reputable architect meeting all the qualifications for practice under sections 4703.01 to 4703.19 of the Revised Code; provided the board may require applicants under this division to furnish satisfactory evidence of knowledge of professional practice and supervision of construction;

A hearing was held before the State Board of Examiners on May 18, 1978. The Board and myself were represented by legal counsel. Please note, prior to the hearing, my Ohio graphic/written design problems, two failures and one passing were destroyed in violation of Section 4743.02 Ohio Revised Code. The Board's Executive Secretary stated in the hearing transcript, that he was not aware of that provision. I am not calling this destruction of evidence an intentional act. However, I would like to note that the design portion of the multiple choice NCARB professional exam is the only section that I have not passed--by two percentage points. Also, per Exhibit B-3, Building Design and Construction, "A/E Licensing Under Attack in Several States," May, 1979, p. 15; ...when a legislative committee had the (Wisconsin) board ask a panel of Wisconsin architects to regrade portions of the December, 1973 NCARB exam.

The results of that regrading correlated badly with the official results, which led the state to eliminate the (graphic) design portion from the state's architectural exam."

Evidence was presented at the May 18, 1978 hearing documenting my six-year Master of Architecture degree and satifactory completion of three years of experience. In regard to the "satisfactory" evidence of knowledge of professional practice and supervision of construction which the Board at their discretion could also require, I called the Board's Executive Secretary one week prior to the hearing in front of my attorney's law clerk to get directions. The Executive Secretary stated that he did not know what would be required per that phone conversation and this also was documented in the hearing transcript.

The Board did not require any evidence of knowledge of professional practice and/or supervision of construction. However, we did present documentation showing, that while a graduate architectural student, I did complete a graduate level architectural professional practice course at The Ohio State University when the regular professor was hospitalized. My duties included completing the remaining lectures, grading midterm and final exams and determining final grades. I would like to note that this course work was completed by myself without other faculty supervision. In regard to knowledge of supervision of construction, an affidavit was submitted by an expert in the building industry in which he described the importance and extra benefits of a combined business/architectural background to supervision of construction activities (see Exhibits C-3 and C-4).

The Board's Executive Secretary, in response to their legal counsel's questioning, testified at the May 18, 1979 hearing, that to his knowledge, no one had ever been admitted to practice under Section 4703.08 (A) O.R.C. and that I was the first person to ever apply (see Exhibit C-1).

Per an investigation conducted in July, 1979, by Senator M. Morris Jackson, Ohio Senate President Pro Tempore, on my behalf regarding the history of all applications approved under section 4703.08 (A) O.R.C., a list of approximately 350 approved applications was documented (see Exhibits C-12 thru C-16). Unfortunately, the above facts could not enter the court as evidence since the Board's denial of certification was upheld by the Court of Common Pleas and was already in the Court of Appeals process.

The Board's Executive Secretary contended to Senator Jackson that in 1946 the rules of the Board were amended to permit a written examination under the provisions of Section 4703.08 (A) O.R.C.

> Section 4703-1-02 Rules--relative to filing applications
>
> (F) In consideration of applications for registration by exemption under Section 4703.08, of the Revised Code, the Board reserves the right to exercise the discretion provided therein, by requiring the applicant to qualify by passing the technical examination, or in lieu of the technical examination the applicant may submit a Certicate from the National Council of Architectural Registration Boards showing satisfactory qualifications which, in the opinion

of the Board, are equal to those required in the State of Ohio at the date of application.

By comparing Section 4703.08 (A) O.R.C. to the above rule it becomes clear that the rule does not give the board the discretion to totally disregard this section of the law.

The Analysis of Senate Bill No. 15, (as passed by the Senate) Ohio's 113th General Assembly--Regular Session 1979-1980, a bill which added further restrictions to Architectural professional entry, paraphrases Section 4703.08 (A) O.R.C. and states that it is one of the two basic methods of establishing proficiency as specified in the law to obtain, architectural certification. However, also documented in this analysis is the following statement: "According to the Board this method was used to "grandfather" into the system architects practicing at the time the law passed, but is not currently used by the Board to grant certification." (See Exhibit C-7). By reviewing exhibit C-8 (a,b,c and d), Section 1334-7 (A) and (C) General Code and amended forms of Section 4703.08 (A) O.R.C. in 1965, 1969 and 1973, it becomes clear that the actual "grandfather" clause, (Sec. 1334-7 (C)), was repealed in 1932 and that Section 4703.08 (A) was still a viable alternative according to the Ohio Legislature at the time I applied.

The Board denied my application under Section 4703.08 (A) O.R.C. with the justification "that the applicant did not provide any evidence at the hearing which would show that he possessed any special qualifications for licensure without taking the examination required under Section 4703.07, Ohio Revised Code" and "that the Applicant at the time of the hearing had not passed the examination required by Section 4703.07, Ohio Revised Code."

In rebutal, the statute only calls for "satisfactory" evidence and the facts presented strongly indicated that the Board had determined prior to the hearing that no matter what evidence I submitted, their willingness to license under the "license without examination statute", (a separate alternative statute for licensure), would be negative.

The Cuyahoga County Court of Common Pleas in November, 1978, upheld the Board's order without justification. Recently, the Cuyahoga County Court of Appeals, Eighth District, which had the opportunity to review the Analysis of Senate Bill No. 15 -- Ohio 113th General Assembly, also upheld the Board's order with in my opinion the following major justification.

> the Board did not abuse its discretion"...with respect to the furnishing of" satisfactory evidence of knowledge of professional practice and supervision of construction", we find the Board of Examiners from the evidence presented in the matter sub judice could have, and we believe did find the evidence submitted was insufficient to demonstrate satisfactory evidence of "knowledge of professional practice and supervision of construction." Against the background of such a finding, the issue of abuse of discretion does not arise.

> Assuming arguendo that this presents an issue with respect to abuse of discretion, we note that the term "abuse of discretion" connotes more than an error of law or judgements; it implies an unreasonable, arbitrary, or unconscionable attitude. An action is arbitrary if it is without some fair and substantial cause or reason, or if it has no evidence to support it."

The presented facts including the analysis of Ohio Senate Bill No. 15 indicate, in my opinion, that an abuse of discretion did occur. It is most difficult to spend over one year in administrative appeal proceedings with the conceptual framework of the matter predetermined before the presentation of facts eliminating probability of true due process of law. One of my major justifications for presenting my personal case analysis here is to illustrate the need for this Congress to design a standardized due process conceptual framework of professional testing guidelines to ensure that all matters of business handled by state licensing boards provide for a continuous process of due process.

In regard to the critiques of NCARB professional exam per se, I will briefly present information from relevent research data accumulated.

"An Evaluation of the Examinations Administered by the Architects Section of the Architects, Professional Engineers, Designers and Land Surveyors Examining Board," submitted to the State of Wisconsin, Department of Regulation and Licensing by Dr. Glenn E. Tagatz of International Personal Services, Inc., Milwaukee, Wisconsin, in February, 1976, made the following findings: The NCARB examinations do not meet the specifications of the American Psycological Association (APA) with regard to use of psycological measurement technology in exam preparation with respect to job skills; nor the Equal Employment Opportunity Commissions's (EEOC) guidelines for job relatedness; nor the Uniform Guidelines on Employee Selection Procedures of the Equal Employment Opportunity Coordinating Council (EEOCC) (See Exhibits B-1 and A-3).

The "Content Analysis of the NCARB Examination" (Exhibit B-2), prepared for the State of Wisconsin, Department of Regulation and Licensing, prepared by Dr. Peter E. Schriber, Mr. Michael Priestley and Mr. Brent Cambell of National Evaluation Systems, Amherst, Massachusetts, in June, 1979, concluded the following:

> "The most important conclusion is that the match between NCARB examination content and the Wisconsin Task List was less than 50% on an item-by-item match basis. Also, the proportionality of the NCARB examinations differed considerably from that of the Task List." Please note that "EEOC Guidelines state that test content should be as close to 100% job-related as practical and that the further away from 100% the less job-related the test is."

I do not want to imply that NCARB's use of its current exam is an intentional act to restrain trade. However, an only 50% job related exam may in effect do so. The exam is graded on the basis of relative distribution of scores. The national pass rate is approximately 54% more or less each year. Per the 1978 NCARB Architectural Registration Handbook, National Council of Architectural Registration Boards and

Architectural Record Books, 1978, p. 14, "All of the score distributions for the Professional examination have relatively small standard deviations, that is, the dispersion of scores is only about one-half of what might be expected in tests of this length. This would suggest that the examinees who actually took the test tended to be relatively homogenous (i.e., quite similar in terms of both personal characteristics and ability), an assumption which is borne out by the description of candidates in terms of age, educational level, etc."

Since the exam is less than 50% job related, and score distributions have relatively small standard deviations, guessing well on exam questions seems to be the primary vehicle into the profession (based on NCARB's allegation that test candidates are quite similar in terms of ability).

Another problem must be presented concerning the lack of uniformity in applicant "intern" education. The "intern" period ranges from one to five years depending on college degree status, (thirteen years in some states for non-degree candidates), with three years being the average for five-year professional architectural degree holders (Bachelor of Architecture). The problem concerns the diversity of architectural practices in both their size and the building type markets they are involved in. Many "intern" architects are only exposed to work areas that return maximum profits to their employers. Without a required uniform "intern" training program, incentives will continue to exist which motivate employers to exploit this captive "cheaper-labor" "intern" labor force rather than providing the kinds of training these employees are paying for in reduced wages and in effect increase their chances (assuming a valid exam is given) of passing their architectural licensing exam (s).

Mr. Patrick Clinton's article "A/E licensing under attack in several states," published in the May, 1979 issue of Building Design and Construction, pp. 14-15, 16, provides an excellent summary of current NCARB validity problems across the country (see Exhibit B-3).

Proform's article "Wisconsin Architects Balk at National Exam--Design Their Own," published by the National Center for the Study of Professions, Washington, D.C., November, 1978, p. 3, gives an excellent account of activities in Wisconsin leading up to NCARB examination elimination and replacement (Exhibit B-5).

Mr. Ron Shattil's article, "A Critique of the Registration Examination as a Measurement of Beliefs Rather than Ability," American Institute of Architects Journal, November, 1976, pp. 78-79, 100 and letters to NCARB from concerned parties (Exhibits B-6 and B-7) give good accounts of content problems associated with the NCARB exam.

Illustrations of recent concerns expressed by Federal Agencies regarding the negative economic effects of our increasingly expanding professional cartels are found in the following statements:

> "An Address by Michael Pertschuk, Chairman, Federal Trade Commission, before the American Enterprise Institute Licensure Conference", Washington, D.C. (Exhibit A-1) Feb. 22, 1979 and

"Occupational Licensing: Consumer Protection or Cartel Management?", remarks by Mr. Donald L. Flexner, Deputy Attorney General, Antitrust Division, Department of Justice, before the Kentucky Legislative Research Commission's Program on Occupational Licensing, Lexington, Kentucky, February 2, 1979 (Exhibit A-2).

The statements made by Mr. Pertschuk and Mr. Flexner illustrate our need to be concerned about professional cartels limiting competition to keep prices high. Mr. Pertschuk stated one of the problems is the state licensing of professionals, a process usually dominated by members of the professions who hold down the number of entrants to control competition.

I have been cooperating with the Federal Trade Commission in an investigation of the NCARB exam since March, 1979. Also, I have been following, with much interest, efforts by professional political action committees aimed at inhibiting the FTC's regulatory authority which is directed towards maintaining a competitive economic system that does not unreasonably constrain private economic activities. I am hopeful that this Congress will carefully analyze the trade offs of their actions toward the FTC with respect to the balance of our mixed system of government and private enterprise as well as our mixed system of monopoly and competition.

A major issue to be evaluated is the direction of our administrative legal system. Unlike our civil and criminal systems, which have developed with us over the centuries, the administrative system was born early in this century. Unchecked, it could have strong negative effects on our economic and political system. I pray that our American Economic Dream of growth and properity continues for centuries into the future. Ensuring that professional entry is maintained at reasonably competitive levels in consideration of providing for the public health, safety and welfare is but one component of the chain of events that will insure the American Dream.

EXHIBITS

-A- FEDERAL AGENCY CONCERNS

A-1: Michael Pertschuk, Chairman, Federal Trade Commission, Wash., D.C., An Address before the American Enterprise Institute Licensure Conference, Feb. 22, 1979.

A-2: Donald L. Flexner, Deputy Assistant Attorney General, Anti-trust Division, Dept. of Justice, Wash., D.C., "OCCUPATIONAL LICENSING: CONSUMER PROTECTION OR CARTEL MANAGEMENT?" before the Kentucky Legislative Research Commission's Program on Occupational Liçensing, Feb., 1979.

A-3: "EEOC MAY TAKE ANOTHER LOOK AT TITLE VII JURISDICTION OVER LICENSING AGENCIES," Proforum, published by the National Center for the Study of Professions, Wash., D.C., April, 1979, pp. 1-2.

-B- CRITIQUES OF THE NATIONAL COUNCIL OF ARCHITECTURAL REGISTRATION BOARDS (NCARB) EXAM -- written/implemented 1972.

B-1: Dr. Glenn E. Tagatz of International Personnel Services, Inc., Milwaukee, Wisc., "AN EVALUATION OF THE EXAMINATION ADMINISTERED BY THE ARCHITECT'S SECTION OF THE ARCHITECTS, DESIGNERS AND LAND SURVEYORS EXAMINATION BOARD," Submitted to: State of Wisc., Dept. of Regulation and Licensing, Madison, Wisc., Feb., 1976.

B-2: Dr. Peter E. Schriber, Michael Priestley and Brent Cambell of National Evaluation Systems Inc., Amherst, Mass., "WISCONSIN ARCHITECTS LICENSING PROJECT, CONTENT ANALYSIS OF THE NCARB EXAMINATION -- FINAL REPORT," prepared for: Dept. of Regulation and Licensing, State of Wisc., Madison, Wisc., June, 1979.

B-3: Mr. Patrick Clinton, Contributing Editor, Building Design and Construction. "A/E LICENSING UNDER ATTACK IN SEVERAL STATES," May, 1979, pp. 14-15, 17.

B-4: Mr. Oliver Witte, Editor, Building Design and Construction. "LEGISLATURES ASK: WHY LICENSE A/E's?" May, 1979, p. 9.

B-5: "WISCONSIN ARCHITECTS BALK AT NATIONAL EXAM -- DESIGN THEIR OWN," Proforum, published by the National Center for the Study of Professions, Nov., 1978, p.3.

B-6: Mr. Richard C. Them, President, Them Assoc., Inc., Waupaca, Wisc., letter to Dr. William Gorth, PH. D., National Evaluation Systems, Amherst, Mass., re: NCARB exam, Dec. 7, 1977.

B-7: Mr. Thomas Vonier, Project Director, A.I.A. Research Corp., Wash., D.C., letter to Mr. Samuel T. Balen, Director, NCARB, re: NCARB exam, Jan. 3, 1978.

-C- CASE HISTORY OF MR. SHELDON R. GOLDSTINE -- APPLICATION TO PRACTICE ARCHITECTURE.

C-1: "ONE ARCHTECT IN SEARCH OF A LICENSE," Proforum, published by the National Center for the Study of Professions, Fe., 1979, pp. 1-3.

C-2: References: Sheldon R. Goldstine, on file with State Board of Examiners of Architects since late 1976. (Ohio Board)

C-3: Sheldon R. Goldstine, under Sec. 4703.08 (A) O.R.C. (1334-7 (A) G.C.), re: knowledge Professional Practice.

C-4: Sheldon R. Goldstine, under Sec. 4704.08 (A) O.R.C. (1334-7 (A) G.C.), re: knowledge Supervision of Construction.

C-5: Mr. Irving Rothenberg, letter to Mr. William N. Wilcox, Exec. Sec'y., State Board of Examiners of Architects/Ohio, re: Sheldon R. Goldstine, Feb. 22, 1979.

C-6: Dr. Barry I. Fireman, Superintendent, Western Reserve Psychiatric Habilitation Center, letter to Mr. William N. Wilcox, re: Sheldon R. Goldstine, Feb. 20, 1979.

C-7: Analysis S.B. 15/ 113th General Assembly/ Ohio, proof that I never received due-process in my application before the Ohio Architect's Board under Sec. 4703.08 (A) O.R.C.

C-8: History Section 4703.08 (A) Ohio Revised Code (Sec. 1334-7 (A) General Code)
Actual "Grandfather's Clause Sec. 1334-7 (C) G.C.
Refer supplements a,b,c and d to review original law plus amendments in 1965, 1969 and 1973.

C-9: Letter from Sheldon R. Goldstine to Mr. William N. Wilcox, Executive Secretary, State Board of Examiners of Architects, State of Ohio, April 22, 1979.

C-10: Letter from Sheldon R. Goldstine to Mr. William N. Wilcox, April 25, 1979.

C-11: Letter from Mr. William N. Wilcox to Sheldon R. Goldstine, April 27, 1979.

C-12: Letter from Honorable M. Morris Jackson, President Pro Tempore, Ohio State Senate, to Mr. William N. Wilcox, re: history of all applications under Section 4703.08 (A) O.R.C. (1334-7 (A) G.C.), July 12, 1979.

C-13: Letter from Mr. William N. Wilcox to Senator M. Morris Jackson, July 1979.

C-14: Letter from Mr. William N. Wilcox to Senator M. Morris Jackson, documenting 350 more or less applications, approved by the Ohio Architect Board under Section 4703.08 (A) O.R.C. (1334-7 (A) G.C.), August 7, 1979.

C-15: Letter from Senator M. Morris Jackson to the Honorable Dean Conley, Representative, Ohio House, re: Section 4703.08 (A) O.R.C.

C-16: Letter from Representative Dean Conley to Senator M. Morris Jackson, re: use of Section 4703.08 (A) O.R.C. by Ohio Architects approximately twenty years ago.

-D- AMERICAN INSTITUTE OF ARCHITECTECTS (A.I.A.) DESIGNS/BUILD TASK FORCE MEETING, MARCH 11, 1974, re: architectual profession's need to become more business oriented if it is to survive--consistant with the business/architectural academic program Sheldon R. Goldstine began in 1967.

Again, thank you for this opportunity to testify before you today.

Respectfully and very truly yours,

Sheldon R. Goldstine

Sheldon R. Goldstine, MBA/M. Arch.

cc: Honorable Sam M. Gibbons
Honorable Ted Weiss

Enclosures

Mr. **Weiss**. Thank you, Mr. Goldstine.

Our last witness is Mr. I. Leon Smith, director, Professional Examination Service.

STATEMENT OF I. LEON SMITH, DIRECTOR, PROFESSIONAL EXAMINATION SERVICE

Dr. **Smith**. Thank you.

Mr. Chairman, I am Dr. I. Leon Smith, director of the division of Professional Examination Service responsible for developing certification and licensure examinations in the social sciences, medical sciences, environmental health, and allied health areas.

Professional Examination Service is a not-for-profit corporation which has been engaged in the assessment of the qualifications of health and health-related professional workers for more than 35 years.

Accordingly, PES as well as many of its clients will be subject to the provisions of H.R. 3564 that specifically relate to occupational testing. It is our interpretation that the language of section 3.3 and section 6(a)5 and section 6(b)4 is sufficiently vague to include and require disclosure of actual test items and answers for every administration of an examination given for licensure or certification.

The major purpose of the testimony I will provide today is threefold:

First, I would like to raise some serious questions and issues concerning the negative impact that disclosure will bring to the credentialing field in general.

Second, I will discuss the inappropriateness of the provisions of section 6(c) regarding the setting of passing points for occupational tests.

The disclosure of items on a current examination will require a constant supply of new items in order to provide secure tests. I believe there is a real issue as to whether and to what degree licensing and certifying bodies could develop new examinations for each administration.

Without these examinations, credentialing agencies will have a difficult time selecting qualified practitioners. This, in turn, will reduce the ability of the agencies to provide public protection and satisfy the public's demand for quality health service.

I have considered a number of factors in making this statement. Over 60 percent of our clients use their own unique examinations as licensing and credentialing criteria in addition to the examinations we developed. H.R. 3564 would require disclosure of these test items as well, whether they be written, oral, multiple choice, essay, practical, performance or demonstration.

Notwithstanding the obvious difficulties associated with disclosing several types of these examinations, the item pools in these unique areas tend to be quite small and would quickly be depleted upon disclosure.

A survey of our clients indicates that examinations may be given as many as 10 times per year and in many cases on an individual or small group basis. There is no question that disclosure would seriously reduce the number of times these examinations could be administered.

This in turn would markedly reduce the opportunity now available in certain fields for multiple State licensing within a reasonable time period. This procedure permits candidates to take all of the unique examinations given by States in the same region of the country over a period of a month or two.

This is an extremely important concept which presently increases employment opportunities and provides for much greater professional mobility than would be possible if the multistage sequencing were eliminated.

Several fields employ what is referred to as multistage licensing procedures. This means that a candidate must pass one level before going on to the next in the sequence. Typically, the sequence is offered several times during the year. Under test disclosure conditions, however, with its attendant requirement for new items and examinations, it may only be possible to administer portions of the sequence once during any given year, thereby creating considerable delay in the licensing process.

In other cases, licensing boards may be hard-pressed to meet their statutory responsibilities which would require the licensure process to be implemented twice a year. I believe it would be fair to say that licensing boards who could not meet their statutory responsibilities would be subject to litigation for preventing, or at the very least, delaying occupational employability.

Disclosure will also have a number of adverse effects of an educational nature. Primary among them is that it will encourage instructional programs to be more closely oriented to the specific items contained on the tests. This type of instruction perverts the intention of licensing and certifying examinations and will ultimately reduce the independent value of educational criteria.

Specific examination items should be viewed as merely a representative sample of a competency domain or subdomain of the required occupational knowledge, and not of vital importance in their own right. Disclosure will unavoidably draw attention to specific items and away from the competency areas from which they are derived.

An assumption underlying disclosure is that a licensing and/or certifying examination can be treated as a learning experience and that candidates can acquire knowledge or competency from seeing what they missed along with the correct answer. This is simply not the case.

In a number of licensing areas and in several States the boards literally have no control over who actually sits for the examination. Candidates with limited background and educational experience can and do take the licensing examinations. I believe it would be seriously misleading to suggest that such a candidate could in fact acquire "competency" from viewing items on a specific examination. Learning answers is not the same as acquiring the broad base of knowledge sampled by the items.

Since disclosure will require a new examination for every administration, increased costs can be anticipated but at levels substantially higher than those predicted for admissions testing.

This is because the costs for some of our client's own examinations will be spread across perhaps as few as 10–15 candidates per year as opposed to thousands as is the case with large aptitude testing programs. Several unfortunate consequences can be anticipated.

While increased costs could be minimized by reducing the length and/or scope, in other words, reducing the items or questions of the new examinations, the psychometric quality of the tests will suffer and sampling of the content areas so vital to competency assessment may be compromised.

Perhaps, more unfortunate, the level of increased costs may force some of our clients to eliminate the implementation of their unique examination criteria which are typically oral, written, essay or practical in nature, and rely exclusively on the PES examination which employs a multiple-choice format. This would result in a limitation in the range of methods and skills that are assessed and through which licensing and certifying decisions are made.

The occupational provisions of H.R. 3564 will be thrust into an already volatile social and political climate involving the mixture of sunset legislation and the uniform guidelines on employment selection procedures. Some areas in the licensure field are currently fighting for survival on one hand through the application of sunset laws, and increased demands for further clarification and validation of procedures and methods on the other in response to the joint guidelines.

While there is certainly a question as to whether the uniform guidelines apply to licensure and certification functions, most of our clients want to be as responsive as possible to the validational issues. Since the uniform guidelines clearly have the weight of law, it would appear that they could serve as one form of "truth in testing."

I believe that many licensing and certifying bodies will have all they can handle to meet these guidelines under their current examination and fee structure.

What will the introduction of H.R. 3564 accomplish in this setting? Licensing agencies will have to produce sufficient new forms of examinations that the disclosure provisions make necessary, each of which is responsive to the validity requirements of the joint guidelines.

The net result may very well overwhelm and incapacitate the credentialing system. Given the current political situation in some States, the likely failure of licensing bodies to meet these considerable requirements at a high standard of accomplishment will simply make termination under existing sunset legislation a more attractive alternative in certain fields.

I believe this would be unfortunate for those who are interested in maintaining the quality of excellence in health and health-related services.

With respect to the issue of determining an appropriate passing point on a licensing or certifying examination continues to be one of the most vexing problems facing the credentialing field.

In the normative approach, a bell-shaped or normal distribution of examination scores is assumed, and the characteristics of the score distribution are used both to set a passing point and to provide a rationale for interpreting scores. This approach is based on the assumption that a specifiable proportion of qualified candidates can be expected to meet minimum standards for performance in the occupation.

For example, if a passing point is set at 1.0 standard deviations below the mean of a normal distribution, 84 percent of the population of candidates are assumed to be qualified for the credential. The normative approach is criticized because the standards set are referenced solely to the distribution of ability in the normative sample of examinees, and do not take into account criteria for acceptable performance in the occupation, to which passing examiinees are admitted.

It is possible that all or none of the candidates are competent to practice, but a normative passing point will always result in a specific, predetermined proportion who will pass or fail the examination.

An appropriate decision strategy in credentialing should include a means of linking decisions to minimum criteria for acceptable performance. The objective is to set a standard such that examinees whose true level of ability is above that standard are credentialed while those whose true level of ability is below the standard are not credentialed.

The criterion-referenced approach to setting a passing point was introduced to accomplish this objective. The criterion-referenced approach is characterized by the use of an absolute performance standard rather than the relative standard based on normative data.

An absolute performance standard is a fixed score which represents minimal acceptable performance on the test, such as 75 percent of the items on the test answered correctly which, incidentally, is the legislatively set passing point used by several States in certain licensure areas.

There are a number of systematic strategies that have been developed in order to set a rational standard on a criterion-referenced test. The procedures fall into three categories and can be labeled "judgemental", "empirical", and "combination". In judgmental methods, data are collected from expert judges for setting standards.

Empirical methods require the collection of examinee response data to aid in the standard-setting process.

Combination methods incorporate judgmental data and empirical data into the standard setting process. There are many factors to consider in selecting one of these methods to determine cutoff scores, importance decisions, time availability, and capability of the individuals in applying the particular method.

Given this background, what can be said concerning section 6(c) of H.R. 3564 which specifies that a passing score cannot be set by reference to the relative distribution of scores of the test subjects? This is what I have referred to previously as the normative method.

First, while the provision does correctly assume that passing should depend on the content of the examination and not the performance of the individuals who happen to take the test, it does not prohibit the use of an absolutely arbitrary and unjustified standard.

Second, it is quite an easy matter to get around the provision by simply converting the group average, norm, to an absolute percentage score which is used as the passing point.

Third, the provision would prohibit a potentially defensible strategy in which the passing point is based on the performance of a small, but acknowledged-to-be-competent subgroup of the population taking the examination.

Fourth, and perhaps most important, the provision cannot guarantee that the more rational procedures previously mentioned will be utilized or that the decision to select from among the three types is defensible.

I am forced to conclude that the passing point provision in H.R. 3564 represents an honest but far too superficial attempt to deal with a highly complex issue.

It is my contention that H.R. 3564, while well meaning, represents a cure whose ultimate consequences may overshadow the disease it supposedly addresses. What is at stake is our country's ability to provide quality health and health-related services. This is why I am concerned that the full range of impact of the testing procedure on our quite diverse credentialing systems be given proper consideration.

Toward that end, I would strongly encourage that other agencies in the occupational testing field be invited to testify at subsequent hearings of this subcommittee. In the final analysis, the populist appeal that both bills have so clearly enjoyed up to this point must give way to the documentation and verification of the benefits to be derived from the passage of these bills. Until such data are available, enactment of the proposed legislation is premature.

[The prepared statement of I. Leon Smith follows:]

TESTIMONY OF DR. I. LEON SMITH
PROFESSIONAL EXAMINATION SERVICE
TO THE SUBCOMMITTEE ON
ELEMENTARY, SECONDARY, AND VOCATIONAL EDUCATION
OF THE COMMITTEE ON EDUCATION AND LABOR
CONGRESS OF THE UNITED STATES
SEPTEMBER 24, 1979

Mr. Chairman, I am Dr. I. Leon Smith, Director of the Division of Professional Examination Service responsible for developing certification and licensure examinations in the social sciences, medical sciences, environmental health, and allied health areas.

Certification refers to functions performed by non-governmental agencies and volunteer associations. The process confers occupational status on individuals so designated but does not legally prevent the practice of an occupation using these skills by people who do not have such a certificate. Licensure requires the demonstration of competence based on statutory requirements and anyone who does not have a license is not authorized to practice. Accordingly, licensure represents a higher level of credentialing and is designed to protect the public against incompetent practitioners.

Professional Examination Service is a not-for-profit corporation which has been engaged in the assessment of the qualifications of health and health-related professional workers for more than 35 years. PES has assisted Federal agencies, state and local health departments and personnel agencies, licensing boards, certifying bodies and educational institutions in examining individual competence to practice effectively in health service systems. Accordingly, PES as well as many of its clients will be subject to the provisions of HR 3564 that specifically relate to occupational testing. It is our interpretation that the language of section 3.3 and section 6(a)5 and section 6(b)4 is sufficiently vague to include and require disclosure of actual test items and answers for every administration of an examination given for licensure or certification.

The major purposes of the testimony I will provide today is threefold; first, I would like to raise some serious questions and issues concerning the negative impact that disclosure will bring to the credentialing field in general. Second, I will discuss the inappropriateness of the provisions of section 6(c) regarding the setting of passing points for occupational tests, and finally, I will point out some of the other more specific shortcomings of the provisions as they relate to occupational testing. I have attempted to limit my presentation to issues or variations of issues that have not been presented in previous testimony. I believe that some of our clients will also be submitting position papers that will provide more specific examples of the general issues that I shall be mentioning.

Potential Negative Effects of Disclosure

The disclosure of items on a current examination will require a constant supply of new items in order to provide a secure test. I believe there is a real issue as to whether and to what degree licensing and certifying bodies could develop new examinations for each administration. Without these examinations, credentialing agencies will have a difficult time selecting qualified practitioners. This, in turn, will reduce the ability of the agencies to provide public protection and satisfy the public's demand for quality health service.

I have considered a number of factors in making the above statement. Over 60% of our clients use their own unique examinations as licensing and credentialing criteria in addition to the PES examination. HR 3564 would require disclosure of these test items as well, whether they be written, oral, multiple-choice, essay, practical, performance or demonstration. Not withstanding the obvious difficulties associated with disclosing several types of these examinations, the item pools in these unique areas tend to be quite small and would quickly be depleted upon disclosure. Public disclosure, then, would seriously weaken and/or eliminate test item banks and would have a significant impact on small but critical areas of knowledge where only a few items might be appropriate.

A spot survey of our clients indicates that examinations may be given as many as 10 times per year and in many cases on an individual or small group basis. There is no question that disclosure would seriously reduce the number of times these examinations would be administered. This would markedly reduce the opportunity now available in certain fields for multiple state licensing within a reasonable time period. This procedure permits candidates to take all of the unique examinations given by states in the same region of the country over a period of a month or two. This is an extremely important concept which presently increases employment opportunities and provides for much greater professional mobility than would be possible if the multi-state sequencing were eliminated.

Several fields employ what is referred to as multi-stage licensing procedures. This means that a candidate must pass one level before going on to the next in the sequence. Typically, the sequence is offered several times during the year. Under test disclosure conditions, however, with its attendant requirement for new items and examinations, it may only be possible to administer portions of the sequence once during any given year, thereby creating a considerable delay in the licensing process. In other cases, licensing boards may be hardpressed to meet their statutory responsibilities which would require the licensure process to be implemented twice a year. I believe it would be fair to say that licensing boards who could not meet their statutory responsibilities would be subject to litigation for preventing, or at the very least delaying occupational employability.

Disclosure will also have a number of adverse effects of an educational nature. Primary among them is that it will encourage instructional programs to be more closely oriented to the *specific* items contained on the tests. This type of instruction perverts the intention of licensing and certifying examinations and will ultimately reduce the independent value of educational credentials. Specific examination items should be viewed as merely a *representative sample* of a competency domain or subdomain of the required knowledge, and *not* of vital importance in their own right. Disclosure will unavoidable draw attention *to* specific items and *away* from the competency area(s) from which they were derived.

An assumption underlying disclosure is that a licensing and/or certifying examination can be treated as a learning experience and that candidates can acquire knowledge or competency from seeing what they missed along with the correct answer. This is simply not the case. In a number of licensing areas and in several states, the boards literally have no control over who actually sits for the examination. Candidates with limited background and educational experience can and do take the licensing examinations. I believe it would be seriously misleading to suggest that such a candidate could in fact acquire "competency" from viewing items on a specific examination. Learning answers is not the same as acquiring the broad base of knowledge sampled by the items.

Since disclosure will require a new examination for every administration, increased costs can be anticipated but at levels substantially higher than those predicted for admissions testing. This is because the costs for some of our client's own examinations will be spread across perhaps as few as 10-15 candidates per year rather than 200,000 or more as is the case with large aptitude testing programs. Several unfortunate consequences can be anticipated. While increased costs could be minimized by reducing the length and/or scope (i.e., items, questions) of the new examinations, the psychometric quality of the tests will suffer and sampling of the content areas so vital to competency assessment may be compromised. Perhaps more unfortunate, the level of increased costs may force some of our clients to eliminate the implementation of their unique examination criteria which are typically oral, written, essay or practical in nature, and rely exclusively on the PES examination which employs a multiple-choice format. This would result in a limitation in the range of methods and skills that are assessed and through which licensing and certifying decisions are made. Even the issue of obtaining the necessary additional financial resources may present a problem to licensing boards whose fee structure in many cases is set by the state legislature.

The occupational provisions of HR 3564 will be thrust into an already volatile social and political climate involving the mixture of sunset legislation and the uniform guidelines on employment selection procedures. Some areas in the licensure field are currently fighting for survival on one hand through the application of sunset laws and increased demands for further clarification and validation of procedures and methods on the other in response to the joint guidelines. While there is certainly a question as to whether the uniform guidelines apply to licensure and certification functions, most of our clients want to be as responsive as possible to the validational issues. Since the uniform guidelines clearly have the weight of law, it would appear that they could serve as one form of "truth in testing". I believe that many licensing and certifying bodies will have all they can handle to meet these guidelines under their current examination structure. What will the introduction of HR 3564 accomplish in this setting? Licensing agencies will have to produce sufficient new forms of examinations that the disclosure provisions make necessary each of which is responsive to the validity requirements of the joint guidelines. The net result may very well overwhelm and incapacitate the credentialing system. Given the current political situation in some states, the likely failure of licensing bodies to meet these considerable requirements at a high standard of accomplishment will simply make termination under existing sunset legislation a more attractive alternative in certain fields. I believe this would be unfortunate for those who are interested in maintaining the quality of excellence in health and health-related services.

Passing Point Procedures

The issue of setting, selecting, or determining an appropriate passing point on a licensing or certifying examination continues to be one of the most vexing problems facing the credentialing field. Two approaches to setting passing points are currently in wide use by boards and certifying agencies. Both of these approaches have shortcomings.

In the normative approach, a bell-shaped or normal distribution of examination scores is assumed, and the characteristics of the score distribution are used both to set a passing point and to provide a rationale for interpreting scores. This approach is based on the assumption that a specifiable proportion of qualified candidates can be expected to meet minimum standards for performance in the occupation. For example, if a passing point is set at 1.0 standard deviations below the mean of a normal distribution, 84% of the population of candidates are assumed to be qualified for the credential. The normative approach is criticized because the standards set are referenced solely to the distribution of ability in the normative sample of examinees, and do not take into account criteria for acceptable performance in the occupation to which passing examinees are admitted. It is possible that all or none of the candidates are competent to practice, but a normative passing point will always result in a specific, predetermined proportion who will pass or fail the examination.

More on this procedure in a moment.

An appropriate decision strategy in credentialing should include a means of linking decisions to minimum criteria for acceptable performance. The objective is to set a standard such that examinees whose true level of ability is above that standard are credentialed while those whose true level of ability is below the standard are not credentialed. The criterion-referenced approach to setting a passing point was introduced to accomplish this objective. The criterion-referenced approach is characterized by the use of an absolute performance standard rather than the relative standard based on normative data.

An absolute performance standard is a fixed score which represents minimal acceptable performance on the test, such as 75% of the items on the test answered correctly, which incidentally is the legislatively set passing point used by several states in certain licensure areas. In this approach, assumptions are made about the characteristics of the test rather than about the characteristics of the examinees. In setting a passing point at, say, 75% on an examination, the inference is made that an individual who responds correctly to 75% of the particular items on a given form of an examination is able to perform at or above the minimum acceptable level in the occupation.

Problems in the use of absolute performance standards arise when they are applied to written, objective examinations which measure complex or abstract domains. Most written credentialing examinations measure the cognitive aspects of occupational performance. The examination samples knowledges and skills across all of the domains comprising the occupation. The score on this examination is then usually used to make an inference about the candidate's overall pass/fail status vis-a-vis entrance into the occupation. Thus, the connection between the test and actual job performance in indirect because the examination assesses the knowledge rather than actual job performance.

In the absence of direct sampling of job behaviors, users of such an examination employ assumptions about acceptable performance in their attempts to establish a defensible passing point. Unless these assumptions are based on some systematic procedure or strategy for connecting test performance with job performance, they tend to be arbitrary and possibly unjustifiable.

There are a number of systematic strategies that have been developed in order to set a rational standard on a criterion-referenced test. The procedures fall into three categories and can be labelled "judgmental", "empirical", and "combination". In judgmental methods, data are collected from judges for setting standards, or judgments are made about the presence of variables (for example, guessing) that would affect the selection of a standard. Empirical methods require the collection of examinee response data to aid in the standard-setting process. Combination methods incorporate judgmental data and empirical data into the standard setting process. There are many factors to consider in selecting one of these methods to determine cut-off scores.

1. How important are the decisions?
2. How much time is available?
3. What resources are available to do the work?
4. How capable are the appropriate individuals in applying a particular method successfully?

Several of these methods are currently in use in the credentialing field and have been supported in litigation.

Given this background, what can be said concerning section 6(c) of HR 3564 which specifies that a passing score can not be set by reference to the relative distribution of scores of the test subjects? This is what I have referred to previously as the normative method. First, while the provision does correctly assume that passing should depend on the content of the examination and not the performance of the individuals who happen to take the test, it does not prohibit the use of an absolutely arbitrary and unjustified standard.

Second, it is quite an easy matter to get around this provision by simply converting the group average (norm) to an absolute percentage score which is used as the passing point. Third, it would prohibit a potentially defensible strategy in which the passing point is based on the performance of a small, but acknowledged-to-be-competent subgroup of the population taking the examination. Fourth, and perhaps most important, it can not guarantee that the more rational procedures previously mentioned will be utilized or that the decision to select from among the three types is defensible. I am forced to conclude that the passing point provision in HR 3564 represents an honest but far too superficial attempt to deal with a highly complex issue.

Specific Shortcomings of HR 3564

I have mentioned the general reasons why HR 3564 is an inappropriate means of addressing the stated or implicit objectives of the "Truth in Testing" Act. This judgment is based in part on a careful study of the bill's provisions and the conclusion that they frequently reflect an incomplete understanding of well established procedures. The following comments are provided on specific provisions in the bill not already covered in my previous remarks.

Section 2(2) and Section 3(2)

Given the evolution of the Supreme Court's interpretation of the Commerce Clause, it appears likely that the proposed Truth in

Testing Act would most likely survive any serious constitutional attack based upon the Commerce Clause. However, I believe that the interstate aspects of _each_ occupation must be individually judged and examined.

Section 3.3, 6(a)(1-4), 6(b)(1-3)

Except for the disclosure issue, licensing and certifying agencies _currently_ make information available to candidates that would satisfy these requirements.

Conclusion

It is my contention that HR 3564, while well-meaning, represents a cure whose ultimate consequences may overshadow the disease it supposedly addresses. What is at stake is our country's ability to provide quality health and health-related service. This is why I am concerned that the full range of impact of the bill on our quite diverse credentialing systems be given proper consideration. It is critical that additional examination of HR 3564 focus on balancing the need for the legislation as up against the impact it could have. The potential adverse effects clearly deserve more study than can be provided here. Toward that end, I would strongly encourage that other agencies in the occupational testing field be invited to testify at subsequent hearings of this Subcommittee. In the final analysis, the populist appeal that both bills have so clearly enjoyed up to this point must give way to the documentation and verification of the benefits to be derived from the passage of these bills. Until such data are available, enactment of the proposed legislation is premature.

Mr. Weiss. Thank you very much.

Before we start the questions, let me indicate that I understand Dr. Berendzen has a board of trustees meeting that he has to leave for very shortly, so we will see if we can have the earlier questions directed to you, Dr. Berendzen.

Let me indicate for the record that Congresswoman Shirley Chisholm, who is a cosponsor of H.R. 4949, had intended to be with the subcommittee this morning, even though she is not a member of the subcommittee or the full committee, but her duties on the Rules Committee preclude her being here with us.

She did ask a brief paragraph be included in the record, and I am pleased to do that.

She says:

> I want to also emphasize that I have no delusions about what this legislation can accomplish.
>
> I did not envision this bill to be a placebo for the problems that confront minorities in the testing process.

With that, Mr. Erdahl?

Mr. Erdahl. I have no questions of this witness, and I will defer.

Mr. Weiss. Mr. Buchanan, do you have a question of Dr. Berendzen?

Mr. Erdahl. I understood he had to leave, but maybe I could ask a general question and he could respond or someone else could. We have had some very good testimony today both from the proponents and the opponents of the bill, and I want to personally thank every member of the panel for being here. It has been helpful for all of us.

Somebody mentioned the concern of being excluded from admission. One of the points, I think, that have been stressed by those that are favoring this bill, is that the testing system is somehow unfair, because it could exclude people from admissions to a profession or to an educational opportunity. We have had other people who are the opponents who have stressed, and properly so, that certainly testing should not be the sole criteria, for admission. I say this somewhat facetiously, but we should be looking at legislation to say that testing only be considered as 50 percent or 10 percent or 9 percent.

I don't think we can really do that but, hopefully, it will be considered as only a part of it.

One of the general questions would be how much emphasis do colleges and universities put on standardized tests in their admissions policies?

Some members of the panel from their knowledge could answer that.

What really is, if there is a relative weight put on the testing, standardized testing, what is it?

Dr. Berendzen. I am Richard Berendzen, and I gather the question is directed to me.

At the American University, and also at some other institutions that I have been at before, you could find it a fairly common practice across the United States, and that is, that the standardized tests are one of perhaps eight to ten different criteria that are evaluated in the review of the applicant.

We can segregate the applications into the undergraduate level and those into the graduate and professional schools, such as a law school, medical school. In all of these situations the standardized tests certainly carry weight, because they provide us with an opportunity to normalize students from a diversity of schools, geographical locales, et cetera, against one another.

On the other hand, it is indeed all of those other issues that must be taken into account by an astute admissions committee and admissions office.

It has been alluded to earlier today that students from different economic backgrounds score with different abilities on those tests, and that certainly is true.

A half century ago Will Rogers said much the same thing, that we are all ignorant, only about different things.

Well, with respect to the standardized tests, we are quite mindful of that, and what we have to insure is that an applicant, say, from an inner city school, is being treated in a just and reasonable way, when we review the total package of the application file.

Now, it should be remembered that application is really a two-way street. It is not merely for the exclusion of students. It is also for the inclusion of appropriately qualified students.

It would be a significant disservice to a student to be accepted to a university or to a professional program if that applicant were ill-qualified so, in summary, the standardized tests are important. They are only one of several criteria, and what concerns me personally, and I believe the various associations that I am representing today, if those bills were in fact enacted, then I fear that pragmatically what would occur is that universities and colleges would be forced to invent their own admissions criteria.

They would be forced to invent their own testing procedures in addition to the nationally standardized ones. In short, we would be returning to practices that were extant many decades ago.

Those kinds of procedures would lead to a random array of testing processes across the country, and each institution would set its own criteria and be applying its own tests. Those tests would not have the degree of validity of careful testing and review that the current standardized tests have.

I believe in the long term that could turn out to be a significant disservice to the minorities and other ones who are in fact sometimes speaking in favor of this particular legislation.

Mr. **Erdahl**. Sir, you bring to mind the fact that these tests are not only to exclude people but also to include them in educational opportunities; I assume that some people who perhaps haven't done as well in school with grades because maybe they have not applied themselves as they should, might do rather well in some entrance or aptitude tests.

Mr. **Berendzen**. We must consider what was the high school or preparatory school of the applicant and how well did that correlate with the student's score on the standardized tests; and the admissions officer must weigh all of those considerations.

Sometimes we find to our delight that students from areas that we would not have normally expected it have scored exceedingly well on these tests nationally, and we have also found that the scores on the SAT's end up being a fairly good predictor of the student's success in the freshman year, and particularly good predictors from minority students.

That is not to say anything about how well the person will succeed in life in general or succeed by the time they are into their junior or senior year, but we do get a fairly good indication of the student's entering abilities.

It is not just for the exclusion of students but counselling and advisement of those students in their freshman year.

If a student comes in and is admitted, but provisionally, we have some indication by means of these tests of which students should be channelled into basic skills programs thereby assisting them in the summer prior to their freshman year, and they succeed far better because of it.

Mr. **Erdahl**. I believe Ms. Simmons wanted to respond.

Mr. **Weiss**. Dr. Berendzen has indicated he wants to leave.

In order to accommodate him, perhaps all of us ought to have a chance to ask him one or two questions. If he wants to stay he is more than welcome. We are really trying to accommodate his schedule at this point.

Dr. **Berendzen**. Thank you, Mr. Chairman.

Mr. **Weiss**. I have really only one question, I guess, or one area of questioning.

You heard Dr. Loewen's comments about the lack of correlation between the capacity of students from rural and minority groups, especially to be able to do the work at some of the major universities and the test scores which they achieved on these standardized tets.

Now, do you have any special involvement or have you had a special background in the area of testing? Are you familiar with the literature in the field? Would you comment on Dr. Loewen's studies and his testimony?

Dr. **Berendzen**. Well, I think what he has found is something which is understood and appreciated by most people in higher education, indeed, that there are reasonably strong correlations between the scores that students achieve on nationally standardized tests and the income of their families.

It is also true that the incomes of their families correlates with many other things, social deprivations in a variety of ways, the likelihood of that individual being a victim of violent crime in the street.

There are many, many correlations which associate themselves with income level in American society. What we basically end up proving is the fairly obvious truism, that the people who come from reasonably affluent families very often tend to do better than people who do not.

What do you do on an admissions committee? You must look beyond those rudimentary dimensions and ask what was in fact the geographic locale of the applicant, what was the family employment of the applicant? What was the interview like? What were the letters of reference like?

What was the total package of the student like, so we do not look simply at one diagnostic. If we did, it would all be simplified at looking at the socioeconomic level.

This speaks to the heart of a very relevant question about both of these bills. What is the real issue here?

Is the real issue to kill the messenger as it were in times of yore where you get a result that you don't like to see and therefore you don't want to have the process?

Is the real issue at hand before us whether or not standardized tests in any form are to be continued in the United States or if they are to be modified or destroyed.

Various people who have testified in behalf of the bills have stated that they are not opposed to standardized testing. Yet in fact what they are proposing would basically rob it of its current potential so significantly that I am fearful that what we would end up with, if these bills were passed, would be a system in the United States in which standardized testing basically would end.

What then would happen is that each institution in 4 or 5 years time would be forced into its own procedures and all of the benefits that can be accrued by looking at the Nation as a whole and weighing the factors subjectively and objectively would be lost.

That, I think, would be a disservice to us all, and in particular to the very people that it presumably had been designed to benefit.

Mr. **Weiss.** In listening to your testimony, and the part of it where you suggested that through voluntary efforts of the institutions themselves and of the academic community, we would have a better way of correcting whatever abuses or problems exist, I think back to the testimony I heard at an earlier hearing, the testimony this subcommittee had from Eleanor Norton, Chairman of the Equal Employment Opportunity Commission, the very end of July, beginning of August.

We were talking about discrimination and affirmative action programs in universities; and she said that she would be stunned if she were to receive—she used stronger language than that—if she were to receive the kind of nonresponses and invasions that her Commission receives from the higher education community, and that the corporations would not dream of being as resistant to affirmative action programs in relation to women and minorities as the higher education community is.

So I must tell you, as one who has heard that testimony and has followed the record, I am not overly optimistic about the willingness of the higher education community to recognize even that there is a problem and wants to take steps to correct what abuses that there may be in the area.

Are there further questions of the doctor?

If not, Dr. Berendzen, thank you very much for your participation. We appreciate your time scheduling, and thank you for adjusting it to be with us.

Dr. **Berendzen.** Thank you very much.

Mr. **Weiss.** Mr. Erdahl?

Mr. **Erdahl.** Thank you, Mr. Chairman.

I was going to let some other people on the panel respond to that same question posed to Dr. Berendzen and respond to the testimony we heard in response to my question.

Ms. Simmons, you gave an indication you wanted to make a comment.

Ms. **Simmons**. NAACP has been concerned about how tests have excluded minorities and, based on that, working with the educational testing service and the college entrance examination board, the educational testing service said to us that they were commissioned to design tests to exclude; and, based on that, it just tended to bolster information we had pulled together across the country to show the impact of standardized tests, culturally biased tests on minorities.

I would like to ask Dr. Franklin who is accompanying me if he could respond also, please?

Mr. **Erdahl**. Mr. Chairman?

Dr. **Franklin**. I would like to at least add to my credentials as a consultant to the NAACP. I am also a member of the National Committee on Testing for the Association of Black Psychologists and chairman of the Graduate School Committee on Expanding Educational Opportunities that is involved with all the admissions of graduate students, as well as for I guess responsibility in my department as a committee, a member of the Admissions Committee for the Doctoral Program in Clinical Psychology.

Certainly, one of the things that must be kept in mind is that the standardized test as used in admissions is very often represented as only one of several criteria; but many of us know, when we get down to the selection process, because so much weight is placed upon the standardization equation of this test, that it equalizes a lot of other criteria used in the selection process.

Consequently, people tend in the admission process to rely more heavily upon the standardized test score as an indicator, as a predictor, and let me footnote that the test score really only predicts a certain aspect of academic performance, and does not take into consideration the motivation and interest and the kinds of support systems that a person can develop to achieve success in either school or in a professional occupation.

So that if a person is excluded from a college opportunity based on a college score, he is also being excluded on his ability to develop resources to overcome that score, so it's a misrepresentation, and certainly it is a major issue.

Mr. **Erdahl**. Could I follow up a statement of Ms. Simmons with a question?

Unless I misunderstood you, you give the indication that there has been a deliberateness in setting up a testing emphasis that is discriminatory.

To my knowledge, this is the first time that has come up as a deliberate one. Maybe it was inadvertent because of socioeconomic factors, experiences of some people in communities or rural areas that might not know about Sake or Chianti.

Could you expand on that a bit on who says that? That is obviously a very serious charge and one that this committee should be aware of.

Ms. **Simmons**. I would be very pleased to do that.

It is during the course of our trying to pull together the persons and the issues to be considered by the NAACP's conference on minority testing, and we had at that time Dr. Franklin and some other psychologists, the college entrance examination people, educational testing service people and college board people.

We raised questions with reference to the impact that those tests would have, and we specifically mentioned, as I recall my mentioning LSAT's impact, I was concerned about LSAT's because I am an attorney.

The representative from the educational testing service said: "Look, we were commissioned to design an instrument to exclude. We did."

Dr. **Loewen**. I would like to respond, if I could, to Dr. Berendzen's comments; and I wish he were still here so he could hear my response.

He said, I believe, that the standardized testing is only 1 of 8 or 10 factors that he considered and went on to say that it's the factor that normalizes students from diverse backgrounds.

Since that is entirely contrary to my experience and his testimony, that is precisely what the standardized test fails to do, and the very fact that a responsible educator can state that it normalizes students from diverse backgrounds is a misuse of those tests.

That statement should never have been made by a representative of a testing organization, and it has been made so many times it has now become a cliche, and it is not at all based on any data that I know of whatsoever and I would challenge the Doctor to base it on some.

He stated that the tests are a good predictor of majority and particularly of minority performance in school. I don't believe that for a moment. The predictions that are published by the Miller Analogies test people, ACT, or the ETS, the correlations are in the neighborhood of 0.2 to 0.4, usually 0.3.

That means that the test relationships are being explained 10 percent, because if you square that 0.33, 10 percent of all the variations in first-year grades can be explained by its relationship to the test score.

The correlation drops then for second- and third- and fourth-year grades. The very best correlation is with first-year grades.

Furthermore, if you look from minority to majority you see a vast gap in test score differences, standard deviations on the tests. It's 135 points, let's say, roughly on the SAT.

The black students, compared to white, are not doing that badly in school. I have data in the things that I submitted in my appendix that indicates that black students do almost as well as white students, that they graduate from medical school and graduate from advanced and difficult undergraduate programs at almost exactly the same rate as whites, even when they are admitted with 100 to 200 points lower.

There is no such validity of cross groups, and I just don't believe that the standardized tests are valid cross groups.

I think that he is overstating the matter when he claims that this committee or these bills are against testing.

I am supporting a modification of 4949 which, as I see it, can only improve testing.

The assertion that it would destroy testing is not true.

Mr. **Weiss**. OK.

Thank you.

Dr. **Newsom**. Does he still want a response to that question or are you through?

Mr. **Erdahl**. If the rest of the committee will yield, I don't want to dominate the questions, but I would welcome your response.

Dr. **Newsom**. Mine will be very short.

While I agree with the doctor who just left that we are opposed to the passage of this bill at this time, I do not agree with his reasoning; and I agree with my colleague on the right, that my concern, and in disagreement with him, is that it is an executionary technique.

I would hope that this group would reconsider this from the point of view of looking at it more carefully a little bit longer before you pass this particular regulation.

Mr. **Weiss**. Yes, Dr. Smith?

Dr. **Smith**. Thank you.

In a little different direction, not to muddy the waters too much, much of the discussion is focused on the standardized test in a sense because that is what affects aptitude testing.

To broaden its base a bit, since the Gibbons bill is one of the bills being considered, in my opinion a standardized test is much more general than just an aptitude test. An interview and an oral examination in a sense could be prepared to be a sort of standardized measure.

In many of the health and related fields, laboratory examinations requiring the examination of slides and specimens and so on are a form of specific examination that many licensure bodies use. We don't deal specifically with that kind of test but in a sense, that is a standardized test.

So I think the focus has been perhaps a bit too specific on the implication regarding the standardized test, a la ETS and so on.

In light of that, with respect to the weight given various examinations, various standardized tests, in the licensing area in a broad scale I think it would be fair to say that the importance of practical examination and demonstrations in many health allied areas are not given secondary consideration and indeed are very, very important components of competency.

The examinations that we develop or that any other firm might develop might simply tap some of the cognitive knowledge, the specific knowledge that might be necessary, but these other examinations which I think would seriously be compromised by the passage of the bill, those examinations are very important and I think are a very important part of the weighting process. They are not considered secondary.

Mr. **Weiss**. You are referring to 3563?

Dr. **Smith**. Yes.

Mr. **Weiss**. Let me say that Mr. Miller of California, who is a cosponsor of the legislation, has asked that the record include a series of statements and articles from various newspapers and publications across the country.

Without objection I will offer that for admission.

[The statements and articles referred to above follow:]

[From the Chicago Sun-Times, July 18, 1979]

DEMYSTIFY COLLEGE ENTRANCE EXAMS

College entrance exams have replaced the draft as perhaps the most important and mysterious rite of passage for young people.

672

They can affect the rest of a student's life by determining whether he gets into the college or professional school of his choice.

But after a student takes one of the tests, he's never told which of his answers were wrong. And he is asked to believe, pretty much on faith, that the tests really can predict academic performance.

New York Gov. Hugh Carey recently signed a "truth-in-testing" bill designed to remove some of the mystery from standardized admissions tests. The law allows students to see their graded tests and the correct answers, so they can check for scoring errors and, more important, pinpoint their weak spots.

It also requires testing services to disclose their research into the validity of the tests and to tell students what the tests measure, how the results are interpreted and how to challenge suspected grading errors.

This new law may increase the cost of a test, but we think that will be worth it if it makes the tests more fair and eliminates some of their mystique. And we hope it forces admissions officers to justify their reliance on the tests, which have been accused of being culturally biased and of failing to measure such important traits as creativity, artistic sense and leadership.

The law, of course, applies only to New York. The Illinois General Assembly should consider passing one of its own.

[From the Boston Globe, July 16, 1979]

OPEN TEST LAW APPLAUDED HERE

ANRIG WOULD SPONSOR IT IN MASSACHUSETTS

(By Richard E. Gordon, Globe Correspondent)

Massachusetts Commissioner of Education Gregory Anrig said yesterday that he "reacted very favorably" to a new New York law that requires the results of entrance examinations for college and graduate schools to be opened to public scrutiny for the first time.

The law, signed Friday by New York Gov. Hugh Carey, requires companies administering such tests as the Scholastic Aptitude Test, the Law School Admission Test and the Medical College Admission Test to file copies of the tests and correct answer sheets with the New York State Department of Education. It also requires the firms to give a graded answer sheet to any student who requests it.

"Decisions that affect a persons's career or life can be made by these tests," Anrig said in a telephone interview. "Such matters should be subject to public scrutiny."

Anrig said he would support similar legislation in Massachusetts, but added that it might not be necessary because once the testing companies devise procedures to comply with the New York law, they would find it convenient to do the same in other states.

MIT engineering professor Judah Schwartz, director of Project TORQUE, a Newton-based research group trying to devise alternatives to standardized tests, agreed with Anrig.

"Once the testing companies have, in a state as big as New York, devised alternative procedures, they might just as well do it elsewhere," Schwartz said.

The companies that administer standardized tests have lobbied against the New York bill charging that it would drastically increase the cost of the tests and limit the firms' ability to offer special tests to handicapped persons or those who prefer not to take the test on certain days for religious reasons.

Schwartz, however, said the New York law will have national impact on improving teaching, reducing cultural bias in the tests and making standardized tests less important.

"I think that tests whose content cannot be made public are a frightening thing in a democratic society," Schwartz said. "Standardized tests are part of a funny technology that says all of the world can be reduced to a series of multiple choice questions. That's foolishness.

"Unless one makes the questions public, there is no way to improve the quality of the system," Schwartz continued. "If you have crappy tests and no one sees the questions, they'll continue to be crappy."

[From the New Republic, February 5, 1977]

ETS's Star Chamber

WHY YOUR DAUGHTER DIDN'T GET INTO LAW SCHOOL

The Educational Testing Service of Princeton, New Jersey, rapidly is becoming the cradle-to-grave arbiter of social and economic mobility in America. Its tests determine who will be admitted to the nation's prestigious private schools, to almost all colleges and graduate schools, and even to the practice of medicine and law in many states. The typical educated American breaks the seal of his or her exam book (using the eraser end of the number-two pencil) while still in elementary school, and still is filling in between the pale blue lines well into middle adulthood.

Anyone who must deal with it is completely at the mercy of this private monopoly. He or she must take the tests ETS offers, must supply the information ETS requests, and must pay the fees ETS charges.

Most important, individuals whose lives are crucially affected have no way of checking the accuracy of information ETS supplies to schools, universities and state accreditation boards. ETS has been treated as exempt from the requirements of the Family Education Rights and Privacy Act of 1974, commonly known as the Buckley Amendment. This law requires schools to give students access to files containing information about them and limits the use schools can make of this information without a student's permission. The Privacy Act also created a Privacy Commission to study the effects of the act and make recommendations to the office of HEW that administers it. The Privacy Commission currently is considering whether to recommend that ETS lose it exemption. The following cautionary tale may aid the commission in its deliberations.

The story involves the Law School Admissions Test, one of ETS's most popular offerings. When a candidate applies to take the LSAT, the application form asks whether he or she has taken the test before. This question appears among a series of general clerical questions (name, address, and so on). Applicants have no way of knowing that failure to indicate that they have taken the test before (whether by answering the question incorrectly or merely leaving it blank) will have any significant effect on their final applications.

The reason for asking the question was not clerical. Until last year, the test score report form sent to law schools contained—unbeknownst to the student who was paying to have this information sent—a space labeled "unacknowledged repeater." When this space was marked by an asterisk, it indicated that a candidate applying to take the test for the second time had not indicated that this was the case. The asterisk appeared even if a candidate (possibly by accident) failed to answer the question at all. Ralph Smith, a professor and admissions committee member at the University of Pennsylvania Law School, explained that the "unacknowledged repeater" designation raised" a question of [the applicant's] integrity." A number of other law school admissions officers confirmed that this was the construction admissions committees put on the asterisk.

Smith and several others think the appearance of the asterisk kept some applicants out of law school. Of course there are such a variety of reasons and admissions committee can give for denying admission to any one applicant that it is all but impossible to prove that anyone was excluded because of being designated an "unacknowledged repeater."

In the 1975-76 applicant year, the dean of admissions at the University of Pennsylvania Law School noticed that more and more of these asterisks were appearing on LSAT score reports. He reported this to ETS and asked whether the designation was correct in all these cases. Shortly thereafter ETS responded with a memo sent to all law school admissions deans, which read in part:

> It has recently become apparent that a problem exists with regard to the flagging of candidates as "unacknowledged repeaters" (an asterisk in the area under "Unacknowledged repeaters" in the summary section of reports). This problem has been caused by several factors, some of which raise policy questions which will be considered next spring. The result for the current year, however is that some candidates who are not "unacknowledged repeaters" have been designated as such.

The memo warned that the designation should be disregarded and explained that "it will not be possible to identify all such candidates" (those wrongly identified as "unacknowledged repeaters"). For 1976-77, ETS has dropped the asterisk.

There are several disturbing things about this incident. First of all, since applicants have had no way of knowing that the "unacknowledged repeaters" designation existed, discovery of this error depended on the fortuitous curiosity of one admissions dean. Second, since applicants are not permitted to see the reports ETS sends to law schools, individuals would have had no way of clearing themselves of false charges even if they had known that the problem existed. Third, even after the error was discovered, ETS kept it quiet and apparently made no attempt either to correct its past mistakes or to inform the applicants who were harmed by them. In short, candidates who were perfectly acceptable may have been denied access to their chosen profession without ever suspecting the reason.

ETS avoided any mention of the "unacknowledged repeater" incident in testimony at a Privacy Commission hearing in November. Robert Solomon, vice president of ETS, was asked specifically and repeatedly about errors that might be correctable if students had access to their ETS files, but he failed to mention the asterisk episode and insisted that ETS errors were "to the best of our experience" no problem. Solomon now says the failure to mention it was an oversight: "We had no intention not to mention that." Tom White, an ETS official who deals with law schools, says the episode "just wasn't important enough to mention. All of the discussions at the commission were based on what our system is now doing. No effort to describe the system as it has grown and changed was made."

ETS argues it should continue to be exempt from the Buckley Amendment because of the time and expense that would be required to bring its files into compliance, and because the organizations that contract for its services provide sufficient regulation. Of course far poorer organizations, and others (such as state universities) with many more people looking over their shoulders, have been forced to comply. Even apart from the Buckley Amendment, ETS is far too immune from public regulation and scrutiny, given the function it serves and the power it enjoys. There is no good reason to continue this exemption as well.

Kim Masters is a Washington education writer.

KIM MASTERS.

[From the New York Times, July 5, 1979]

ISSUE AND DEBATE—TRUTH-IN-TESTING MEASURE FOCUSING ON A KEY INDUSTRY

(By Edward B. Fiske, the New York Times)

Last month the New York Legislature passed a "truth-in-testing" bill that would require concerns that give such standardized tests as the Scholastic Aptitude Test to make the questions and correct answers available to students who take them.

The bill, which would apply only to the Preliminary Scholastic Aptitude Test and to tests used for entrance to college and to law, medical and other graduate schools, would also require the Educational Testing Service and similar companies to inform students about the purposes of the tests, what they purport to measure and the methods used to calculate scores. The companies would also be obligated to make public any studies of the tests thay have made, including those that show correlations between scores and such things as parental income.

The issue of whether and how to regulate the multimillion-dollar testing industry is becoming a familiar one. Bills similar to the one passed in New York have been offered, but not passed, in Colorado, Florida, Maryland, Ohio and Texas. California has a new law that requires testing companies to make available copis of complete tests that are similar to those that students actually take.

The New York bill was actively lobbied for by a broad range of consumer, teacher, parent and civil rights groups, including the New York Public Interest Research Group, the Consumer Protection Board, New York State United Teachers and the Parent Teachers Association. It was opposed by college spokesmen, the State Education Commissioner, Gordon M. Ambach, and by the testing industry, which also waged a vigorous lobbying effort.

Early this month the bill is to go to Governor Carey, who is under considerable pressure to veto it. His chief aide on higher educational matters, Henrik K. Dullea, is meeting with proponents of both sides to help the Governor make a decision, which will have important consequences for colleges and universities throughout the state.

Background

The "truth-in-testing" measure was the latest in a series of efforts to apply to the field of education the buyer-seller concept used in commerce.

Plaintiffs in several court cases in recent years have accused schools of "malpractice." They included a Copiague, N.Y., youth who unsuccessfully sued the local school system because he was graduated from high school while unable to read above a third-grade level. The Federal Trade Commission has approved new and stiffer regulations for proprietary vocational schools that require, among other things, pro rata refunds for students who drop out.

The new bills regarding testing are essentially an extension into education who took them each year, it should be subject to greater public scrutiny. "It is supposedly a nonprofit enterprise that is operating under the rules of disclosure of a private company," said Allan Nairn, who is directing a study of the Educational Testing Service on behalf of the consumer activist Ralph Nader. While testing companies have taken steps in recent years to provide consumers with more information about the tests, "they are publishing only what they choose to disclose," he said.

Backers of the bill argued that providing students with the questions they got right or wrong would enable them to find scoring errors, as well as to gain helpful information about academic areas in which the students may be weak. The Puerto Rican Legal Defense and Education Fund argued that the measure would also help expose "biases" in tests that "historically had a discriminatory impact on Puerto Ricans and blacks."

A recent Federal Trade Commission study concluded that some students could improve their scores on the S.A.T.'s and similar tests by going to private "coaching" schools has provided another argument. In testimony on the bill, Marjorie Dunbar, education chairman for the state chapter of the National Association for the Advancement of Colored People, declared: "Many students cannot afford these expensive coaching courses and are therefore denied equal access and opportunity. This bill would enable all students to get access to post-test questions, not just those who can afford the courses offered by 'cram' schools."

Finally, the proponents said that publishing actual test questions would improve the quality of aptitude testing by making them subject to scrutiny by experts and the public.

The supporters rejected arguments that "truth-in-testing" would necessarily increase the cost of tests to students. "By E.T.S.'s own figures, they spend 5 percent of test revenues on new question development but make 22 to 27 percent profit," said Mr. Nairn. They also dismiss the danger that E.T.S., as a nonprofit service organization, would be able to discriminate against New Yorkers either in charges or service if requirements here were more strict than elsewhere.

Against disclosure

Robert J. Solomon, the executive vice president of the testing service, said that his company was "very much in sympathy in a general way with the objectives that the originators of the bill seek to serve." He noted that in recent years the company had substantially increased the amount of information given to students regarding the purposes of the tests, the margin of error and the methods of scoring. For the last two years, he said, free sample copies of old tests had been made ??? approved new and stiffer regulations for proprietary vocational schools that require, among other things, pro rata refunds for students who drop out.

The new bills regarding testing are essentially an extention into education of the "truth-in-packaging" concept that already applies to products on grocery store shelves. "This is just an information bill," said Senator Kenneth P. LaValle, chairman of the Higher Education Committee, one of the measure's sponsors. "We are asking that the companies that prepare these tests give us a little more information about them."

The bills have also come out of a context of greater public concern about standardized tests. The National Education Association, the country's largest teacher group, has been engaged in a major campaign against what it regards as undue reliance and widespread misuse of standardized tests at the elementary and secondary school levels. Students are becoming increasingly sophisticated about the margins of errors in testing, and several instances of mistakes have attracted national attention. Last spring, for example, scores on the Medical College Admission Test were reissued after a computer error was found.

For disclosure

Proponents of the bill began their case with the assertion that, since college and graduate school testing was a major industry with enormous influence over the 360,000 New Yorkers many increased the amount of information given to students regarding the purposes of the tests, the margin of error and the methods of scoring. For the last two years, he said, free sample copies of old tests had been made available to students.

A major objection to the bill revolved around the issue of "equating," or the standard practice of using a few of the same questions on successive tests to develop a base for comparing the scores of students who took the tests at various times. On the S.A.T. tests the "equating" questions were not used in calculating a student's score, as they had been on other tests.

The new legislation had been amended to require disclosure only of scored items, but Mr. Solomon argued that, because on some tests the "equating" questions were also scored, this would still require the testmakers to "totally revise the system."

This and other factors, opponents said, would substantially increase the cost of testing. In his testimony, Mr. Ambach, argued against the legislation on the ground that "publishers will be unable to reuse test questions." Moreover, he said, test development would become more expensive because "many questions are pretested as part of the regular examination procedure."

The testing service estimated that, under the new legislation, the cost of the S.A.T. test, now $8.25, would increase by as much as $5. "While it may appear desirable to have such disclosure for those who want to see their papers and results," said Mr. Ambach, "the cost to all of those tested because of the disclosure provision would outweigh the benefits." The College Entrance Examination Board, for whom the service produces the test, said that its annual profits were less than 3 percent a year, not the "22 to 27 percent" asserted by critics.

Opponents also argued that entirely new tests would reduce "service" to students, because the number of times the tests were administered would have to be cut and the 30-day disclosure period would jeopardize special services to handicapped students. Mr. Solomon also said that this would affect quality as well. "There is a tremendous danger of poorer test questions if we have to come up with 21 forms of a test instead of the current seven," he said.

Proponents also maintained that the incidence of scoring errors that might be caught was minuscule. "In 1976–77 we had requests to hand-score 1081 of the 120,000 law boards and found only four errors, all of them because students put the answers in the wrong spaces," he said.

The outlook

With the Commissioner of Education and many college spokesmen vigorously opposed to the truth-in-testing" bill, it is possible that Governor Carey will decide to veto it once it reaches his desk. On the other hand, it also appears that many of the college officials are misinformed about the actual content of the legislation.

Whatever the fate of this particular piece of legislation, however, it would seem certain that the issue will not go away. The problem of responding to growing demands for new forms of public accountability is one that will be with the testing industry for some time to come.

[From the Boston Globe, July 19, 1979]

TESTING THE TEST-MAKERS

We don't know whether anyone has ever tried to add up the number of tests a student who goes to college and beyond takes in a lifetime, but the number is surely staggering. And, oddly, it is only the tests that are most crucial for academic advancement that students never see after they are corrected and it is only these tests—Scholastic Aptitude Tests, Graduate Record Examinations, Law School Admission Tests, Medical College Admissions Tests—for which the "correct" answers are never provided the students.

That, however, may be changing. New York's Gov. Hugh Carey the other day signed legislation that will require the services that administer those tests to file copies of them with the "correct" answers 30 days after they are administered and to provide the graded answer sheets of any student who requests them. Testing services are not likely in the end to boycott New York, although some are threatening to do so. And once they are set up to meet the requirements of the New York law, they may well decide to institute similar policies nationwide.

If that does not occur, the Massachusetts Legislature should consider similar legislation here to increase the pressure for more open testing practices. We live in the world's most test-oriented society. And while colleges and graduate schools may claim that test results are but one measure they employ in admissions, they do clearly matter. Students have a right to confirm, if only to satisfy themselves, that the scores that shape their lives are accurate.

Further, the whole notion of education in our society revolves, should revolve, around an open exhange of ideas and the accumulation of skills and ideas that can lead to wisdom. The administering of tests for which the "correct" answers are never revealed and the students work never returned tends to convert education into a frantic scramble up the greasy pole. It is fundamentally at odds with what the whole process of education ought to be about.

[From the N.Y. Times, July 24, 1979]

TRUTH IN TESTING

The intemperate responses of some academic testers to New York's truth-in-testing law are certain to damage their cause more than the law itself. The associations that administer admission tests for medical and dental schools have threatened to pull out of New York, and the Educational Testing Service, the nation's biggest testing enterprise, has said it might curtail services. Doubtless, the law will require more work on their part, and there will be added costs. But their complaints are overstated. In fact, the law is welcome; it's time to take the mystery out of college testing.

The law signed recently by Governor Carey applies to such examinations as the Scholastic Aptitude Test and the Law School Admission Test that colleges and graduate schools use to rank their applicants. As of Jan. 1, 1980, a copy of any test offered in New York must be filed with the State Department of Education. In addition, data on how a test was constructed and validated must be made public. What irks the testers most is a provision that allows anyone who takes an exam to receive a graded copy. Questions used for research but not counted in scores and specialized achievement tests are exempt.

The testers say the law means an exam given in New York cannot be used again because its contents are likely to be disseminated elsewhere. They contend, too, that the law will create problems in "equating" tests—insuring, for instance, that a 600 verbal score achieved in 1975 is comparable to a 600 in 1979. The testers therefore project drastic increases in cost to maintain the present level of service. And they're unhappy that the law goes into effect so quickly. This complaint may be justified. It takes months to prepare a test and a slight delay may be in order.

Testing services like ETS have been more open about themselves in recent years and the tests have improved; today, most are fairer to minorities than they were ten years ago. But the testers still haven't faced up to other exam limitations or to the frequent misuse of test results. Although testers caution against it, some schools still rely on scores as absolute indicators of ability. ETS, moreover, is surely exaggerating when it contends that the new law might drive up costs by as much as 50 percent and that even then might not be able to uphold testing standards. Development costs are only a small part of total costs passed on the consumer. And some authorities have argued convincingly that at least four techniques already used by the industry would permit "equating" and validating to continue undisturbed—at little or no extra cost.

In any case, there is the matter of elementary fairness. These academic tests help to shape the course of people's lives—their schooling, their careers, the very sense of their own abilities. Students deserve to know how they are being rated and judged. There are freedom-of-information laws, truth-in-lending laws, truth-in-packaging laws. Why not truth in testing as well?

[From the New Republic, August 25, 1979]

EVENING THE SCORE

ETS has good reason to fear New York's new "Truth-in-Testing" law.

In 1980 more than 200,000 New York state teenagers will sharpen their number two pencils and sit down to take the dreaded Scholastic Aptitude Test or SAT— the two-and-a-half-hour standardized test that will determine whether they are admitted to the colleges of their choice. But the 1980 crop of aspiring students will have many advantages over the millions who have taken the tests before them. Under a new New York state law, recently signed by Govrnor Hugh Carey, the students will be well informed about the validity of the tests and their rights as testing consumers. New York's so-called "Truth-in-Testing" law will subject to public scrutiny for the first time the testing agencies that have long had the power to destroy a person's academic and career aspirations in a single morning.

Testing agencies, the virtual gatekeepers of the American elite, have become entrusted with the task of preserving the meritocratic system by means of multiple-choice exams. The undisputed lord of the nation's testers is the Educational Testing Service, a non-profit organization based in Princeton, New Jersey. ETS puts out the

SAT, almost all major admissions exams for colleges, graduate and professional schools, and entrance tests for hundreds of occupations. Despite the enormous power it has over people's lives, and despite many questions that have arisen about the validity of multiple-choice exams in predicting college and career performance, and about racial and economic biases in the tests, ETS has operated since 1947 totally free from government scrutiny of supervision. People who pay to take the tests are powerless to know how the questions were compiled, what the right answers were, or even if their scores have been added up right.

The "Truth-in-Testing" law, which goes into effect next January 1, applies to all tests that are administered to more than 5,000 persons who are applying to college and graduate schools: namely, the SATs, the MCATs (medical boards), the LSATs (law boards), and the GREs (Graduate Record Examinations). The law contains three major provisions. First, test-takers must be told when they register how their scores will be computed; what the tester's contractual obligation to them is, and how test scores have been found to correlate with important background factors such as race, economic class, and special coaching and preparation for the exams. Second, the law requires testing companies to file information and studies on the validity of the exams with the office of the state education commissioner, where the material will be available to independent researchers, curious parents, teachers, and students.

The most dramatic provision of the bill is one that makes public the formerly top-secret questions and correct answers to the exams. Thirty days after the test's results are in, the tester must file copies of the questions and answers with the commissioner. Furthermore, within 30 days of receiving the results, a student is entitled to a copy of his or her graded answer sheet.

This "sunshine"provision, as it is commonly known, provoked ETS to let loose upon every college president and high school principal in New York a torrent of letters, mailgrams, telephone calls, and memos, all threatening the state educational system with dire consequences. Since the law will breach the security of the exams, ETS argues, most questions can't be reused; many more versions of each test will have to be created at a considerable cost. Now, ETS is warning of increases in test fees by as much as 50 percent in 1980. This, their spokespeople bitingly add, will penalize poor people and minorities. ETS is also predicting that it will have to cut back on the number of testing sessions throughout the year, and curtail special sessions for the bed-ridden and those who observe the Jewish Sabbath (most ETS tests are on Saturdays). In short, says the testing agency, the "Truth-in-Testing" law hurts minorities, orthodox Jews, the poor, the sick, and the handicapped. Quite a sympathetic collection.

According to a recent memo by the College Entrance Examination Board, which sponsors the SATs, "truth-in-testing will only advantage—if it advantages anybody at all—aggressive families whose indignation at low reported scores will impose a higher cost on all other families."

ETS and the College Board would have us believe that because a test like the GRE is given 30 times a year and in special sessions, there need to be 20 or more different versions of the test to comply with the new law. But it takes ETS at least 30 days to grade the exams and the bill allows another 30 days before the questions and answers must be released. With a little simple arithmetic, it is clear that testing could operate on a 60-day cycle and only six versions of a test meed be developed each year.

ETS's own internal studies cast doubt on the company's threats about the cost of complying with the law. According to a 1972 breakdown of how fees are spent, less than five percent goes into the writing of the tests. About 25 percent is profit for ETS and its clients—certainly enough to absorb some of the costs of writing new questions. Besides, only a small percentage of the test questions are reused now.

More severe intimidation tactics are being used by the Association of American Medical Colleges, which sponsors the MCATs. This medical school admission test is the only major exam run by ETS's competitor the Iowa-based American College Testing Service. Less than a week after "Truth-in-Testing" was signed into law, the Association announced that it would stop giving the MCATs in New York State at the end of the year, forcing New York's pre-med students to travel out of state for the exam. After this threat, officers of the state Education commission reread the law and concluded that it applies to tests given anywhere for admission to schools in New York state. Dismayed Medical College Association officials may challenge the constitutionality of such an interpretation.

The hysterical reaction of the admissions boards and testing services is actually quite understandable, in light of what is at stake. Truth-in-Testing is not simply a piece of consumer legislation. It is a challenge to the entire system of psychometrics.

The real threat to the testing companies from having to reveal their questions and the correct answers is not that they will have to make up new questions each time. Rather, it is that the tests will be subject to outside scrutiny and evaluation for the first time.

Since World War II the notion of a hand-picked elite has been made obsolete by the computerized world of the multiple-choice exam. Standardized testing gained respectability during the war when the Office of Strategic Services commissioned psychometricians to develop tests that could allocate human resources as quickly as possible. During the great influx of college applicants after the war, the College Board, along with the Carnegie Foundation and American Council on Education, established ETS as a non-profit organization to administer the SAT and develop new tests.

Under the loose supervision of the College Board, ETS rode on top of a wave of testing mania, doubling its revenues almost every five years. It tests everybody from toddlers applying to nursery schools to aspiring CIA agents, stockbrokers, Peace Corps Volunteers, actuaries, gynecologists, and auto mechanics. During the 1960s, college-age males had to take an ETS exam to prove they were worthy of student draft deferment. Today the bar exams in almost all states are in part administered by ETS. Thriving also on millions of dollars of government and foundation research grants, ETS has become an $80-million-a-year operation, comfortably seated on a 400-acre estate with its own swimming pool, hotel, golf course, tennis courts, and private lake. ETS also has its own zip code.

Before the new law, ETS made it extremely clear that what goes on at the 2000-employee Princeton homestead is strictly confidential, particularly in regard to in-house research on the weaknesses of the tests it gives. The possibility of racial and economic biases in the tests has been a sensitive one for many years. A more recent controversy has concerned the effects of coaching on test scores.

Carl Brigham, the developer of the first SAT and the man for whom ETS's Princeton library is named, was best known for his theoretical writings on the superior intellect of White Anglo-Saxon Protestants. Many small-scale studies indicated the SAT, billed as an "aptitude" test, actually tests acquired skills and knowledge, discriminating against the economically disadvantaged. yet complete data on the problem is not made easily available by ETS. An ETS study called *Cultural Bias in Testing: Challenge and Response*—probably the best study of the subject—has been withheld from outsiders.

On the recent problem of the effects of coaching, ETS has taken even more drastic efforts to thwart the disclosure of its research. ETS still insists that coaching cannot significantly alter test scores, although several internal studies have supported the claim made by the Stanley Kaplan coaching schools that for $275 a student can boost his or her "aptitude" by more than 100 points (out of 600) on the SAT. Lewis Pike, author of an ETS study suggesting that the SAT is susceptible to coaching, was fired last year after 14 years with ETS when the research results he was producing proved too controversial. Obviously the suggestion that coaching can improve test scores undermines the theory that the SAT is an "aptitude" test. It also implies a large advantage to those who can afford coaching.

The ETS company line on the new law's provisions requiring this kind of information to be made available to the public is that it is already available. this is patently false. ETS claims that it merely limits access to "responsible researchers." But, to cite one of many examples that disprove that claim, when an attorney at the FTC's Boston regional office launched an investigation of the Stanley Kaplan coaching schools two years ago, the FTC was forced to subpoena the data it needed from ETS.

To temper its image as an Orwellian monster of the computer testing age, ETS has come down hard on abuses by its clients. When the state of South Carolina used an ETS test to set teachers' salaries, ETS terminated its contract to administer the test in South Carolina. ETS also tries to discourage test-crazed admissions deans from setting arbitrary cut-off scores under which applicants are not considered, and from reading finer distinctions into the scores than the tests are able to make. ETS President William Trumbell wrote in the 1978 annual report, "Testing has suffered more from the excessive expectations of its most devoted advocates than from attacks of its critics."

Yet ETS does little to correct the many testing casualties that students suffer at the hands of its clients, or by ETS's own errors. A student who suspects ETS's computers have made a scoring error can request a hand-score, but no one outside of ETS can check the organizations's figures. And there is no real recourse if ETS sends inaccurate information to a college about a student, because most likely the student will never find out about it. One such scandal became public during the 1975-76 academic year ("ETS's Star Chamber," *TNR*, February 5, 1977). Some of the

680

LSAT score reports sent to law schools contained an asterisk indicating that the applicant was an "unacknowledged repeater"—someone who had failed to indicate on the test registration form that he or she had taken the test before. Applicants were not told that overlooking the seemingly innocuous question about past tests risked incurring the deadly asterisk. Even worse, the ETS system accidentally labeled an unknown number of test-takers incorrectly as "unacknowledged repeaters." ETS eventually notified admissions deans to ignore the asterisk. But it made no effort to rectify the damage already done or to inform the victims of the error why they may have been rejected from law school.

The New York law may lead to similar federal legislation. Last month Representative Ted Weiss of New York, along with Shirley Chisholm of New York, and George Miller of California, introduced a "Truth-in-Testing" bill in the U.S. House. The third bill of this type to be introduced over the past few years, the Weiss bill resembles the New York state law, but requires the testing companies to disclose more financial data as well as their contracts with the test sponsors.

ETS and the College Board are taking the new Weiss bill very seriously, if their concern can be measured by the volume of mail being sent to members of Congress. As in New York, ETS and its proponents are repeating their chilling descriptions of great hikes in test fees, limited access to tests, and the disadvantage of non-aggressive students and families. "But far more costly—in a social sense—would be this bill's injection of the federal government into the college admissions process * * * a process protected by the 1st Amendment to the Constitution," adds one College Board memo sent to members of the House on July 23.

This invocation of the Bill of Rights is hogwash, The Educational Testing Service has a far greater impact on the lives of many young people than the government has, yet it denies them the most fundamental rights of equal protection and due process that the Constitution guarantees to all individuals who are affected by government action. A little sunlight on the nation's most powerful unregulated monopoly is long overdue.

BARBARA DEMICK.

[From the Higher Education Daily, August 29, 1979]

HOWARD UNIVERSITY STUDY CITES LIMITS OF STANDARDIZED TESTS

A higher standardized test score does not necessarily mean a professional school applicant is more qualified for admission, and schools should consider many factors in making admissions decisions, according to a new report of Howard University's Institute for the Study of Educational Policy.

Colleges and universities should recognize the limitations of standardized tests and use them accordingly, while also seeking to improve the tests, according to the study's author, Sylvia Johnson. Johnson, a testing expert, is associate professor of education at Howard and a senior visting fellow at the institute.

Less dependence on tests "will result in greater fairness to individuals, as well as groups, in the long run," Johnson said at a news conference yesterday to announce publication of the Measurement Mystique.

POLICY DECISION

The choice of admissions criteria is a policy decision, not a technical one, according to the study. "The question is, what is that important thing in the future that we want our selection instruments to predict?

"Is it 'ability to survive the first year of medical school?' " "Is it 'lifelong involvement in pure research?' " "Or is it 'personal ego strength and emotional stability sufficient to be involved as a psychotherapist for 40 years?' "

Similarly, "if decision-makers feel that it is useful and important to have fair numbers of Black and female physicians, then a utility function would be chosen for use in selection. This function would give weight to the minority or sex status of applicants," the report said.

Tests might be used to establish minimum competency, but "after you get to a certain point, your chances of success are not really increased by having a higher score," Johnson said yesterday. After that, schools should look at other factors, such as motivation or volunteer activities, she said.

BIAS IN TESTS

Blacks and other minority groups suffer from overdependence on standardized test scores because of test bias, the report says. That is, a test may not accurately measure a given ability for all groups, and "the test scores of Blacks, Chicanos, and other minorities tend to have lower reliabilities than those of whites."

Minorities may not score as well as whites do on standardized exams because of biased questions, lower socioeconomic status, poorer educational preparation, "cold" treatment by examiners or the belief that prejudice will limit their options whatever their test results, according to the study.

Johnson said testers should work to eliminate bias in their exams, but a completely unbiased test is probably not possible "at least in terms of a broad-based test applied all over the country * * *. You simply can't measure minds the way you measure a room."

Johnson praised "truth-in-testing" legislation, such as that enacted in New York and introduced in Congress by Rep. Ted Weiss, D-N.Y. Weiss' bill would require the release of test questions and answers and research on test validity (HED, July 25). Dissemination of that information might "provide greater awareness of the limitations of testing" and "demystify" it, she said.

"An arbitrary test score does not have meaning," Johnson said, pointing out the average medical school admission verbal test score of applicants accepted in 1955–56 is close to the average score of those rejected in 1975–76.

Copies of the report are available free by writing the Institute for the Study of Educational Policy, Howard University, 2900 Van Ness Street NW, Washington, D.C. 20008 or calling (202)686-6686.

[From the N.Y. Times Magazine, September 2, 1979]

THE AMERICAN WAY OF TESTING

MANY FORCES HAVE CONTRIBUTED TO THE DECLINE OF AMERICAN STUDENTS' WRITING ABILITY BUT NONE HAS BEEN MORE EFFECTIVE THAN THE WIDESPREAD USE OF MULTIPLE-CHOICE TESTS

(By Thomas C. Wheeler)

What today inhibits the writing of citizens, scholars and students? Americans, long vigorous and inventive in speech, object to the act of writing. Two out of three abilities expected from education—reading and writing—disintegrate as some insidious-influence, thought by many to be television, captures succeeding generations of children.

In defense of literacy, every wordsmith has his own horror story to tell. In a scrawled note, a bright girl, not long out of high school, wished me a "Mary Christmas." An editor of a national weekly told me she would be wary of hiring a copy editor under 30, for fear of grammatical incompetence. A doctor, judging me physically fit and confiding a woe of his, deplored the inability of medical students he teaches to express themselves clearly in writing. An architect of my generation tells me that bright young architects he has hired cannot put together a coherent report. Although the young resent it, their elders utter more than the usual complaints of an older generation against the language of a new.

"This generation has grown up without learning concentration—you don't develop that by switching channels. What can teachers do but pander to the rapid alteration of mood and attention?" this view, from the chairman of an academic committee on literacy, puts the blame on television.

But an influence as devasting as television is the objective test system, introduced by higher learning—and now used in secondary education. Until the 1950's, students wrote essays in schools because they were expected to write essays on college entrance exams. But the university abandoned the essay requirement by adopting, a generation ago, the entirety objective test for admission, the Scholastic Aptitude Test (S.A.T.).[1]

Before World War II, the objective S.A.T. had been a supplement in written achievement tests in various subjects. But when the S.A.T. became dominant, the achievement tests became objective, too, and also optional. When the university dropped the essay requirement, it failed to recognize the power of the system it launched. Once the college entrance exams were objective, secondary schools asked for less writing. Urged on by test manufacturers, high schools began to use objective tests both to prepare their students and for their own examinations. The university, by sanctioning the objective system, bears a terrible responsibility for the decline of writing in the United States.

---

[1] Most American colleges use the S.A.T.; some, mainly in the South and West, use the American College Test (A.C.T.), developed later, modeled after the S.A.T. and produced by the American College Testing Program in Iowa City.

Today, the test manufacturers—led by the Educational Testing Service of Princeton, N.J.—are profiting from miseducation. E.T.S. produces not only the S.A.T. but a battery of tests used for secondary and graduate education. Although the tests have made writing seem unnecessary—although they have also damaged reading ability—they are too expensive to throw out or replace. Harmful though the tests are, the schools have submitted to a system that runs on its own power. Measuring students means measuring schools—according to the aggregate scores of the enrolled—and few schools dare drop out of the spiral of relative standing. Public schools are also tied to the tests by school boards demanding measurement.

Developed first for the Army in World War I and widely used in World War II to test intelligence and ability, objective tests are gifts of war to civilian life. In the 25 years the S.A.T. has been dominant, American education has been revolutionized. The marketplace has overturned the traditional foundations of learning—reading and writing—more completely than the efforts of any mechanistic theorist. Most Americans are probably tested more than they are taught. Compositions, essay questions, term papers—vigorous thinking—all have yielded to one right answer out of four, to boxes to be checked, blanks to be filled. Objective tests not only carry the prestige of being scientifically accurate—when they aren't—but also provide an easy way of handling the masses by machine. The results might have been predictable to an educational system that valued education. The national system that valued education. The national S.A.T. verbal-aptitude scores have shown a steady decline over a 14-year period. If the scores show anything, they show how poor a teacher objective tests are. The American language—supple, imaginative and alive—has lost ground to the pretense of measurement. "Nobody ever cared about my writing" is a refrain I, as a teacher, have heard in several accents. After two decades of objective education, the "educationally disadvantaged" are not merely the poor and minority groups, but the supposedly well-educated and the well-to-do.

Many colleges now profess doubts about S.A.T. scores and say that in admissions considerations high-school grades and teacher recommendations are more important. One distinguished liberal-arts college, Bowdoin College in Maine, found there was no correlation between high test scores and college performance. Finding the "predictive value of standardized tests" questionable and test scores misleading, it has abandoned the test requirement for admission.

Echoing that point of view, New York's Gov. Hugh Carey signed legislation in July requiring that the public be given access to the tests. "The standardized tests are imprecise and open to potential misinterpretation," the Governor said. "It must be . . . a candidate's right to have access to his results."

Beginning Jan. 1, New York State students will be able to examine both their own tests and the answers. The testing services will also be required to disclose their research on the validity of the tests and to describe what the tests are supposed to measure and how to interpret the scores.

In 1978, for the first time, responding perhaps to criticism of the test, the College Entrance Examination Board released a sample S.A.T. exam to students registering for the College Boards. The test, it is often said, favors middle-class white students at the expense of minority groups. Certainly this is true. But middle-class blacks might do as well on it as their white counterparts. The S.A.T.'s tougher verbal questions can be conquered by informed guesswork, but often the correct answer depends on the sheer luck of the student's reading or experience. In a verbal section, the student is asked to "choose the word or phrase that is most nearly the *opposite* in meaning to the word in capital letters." One example is this:

WHET: (A) expire (B) heat (C) delay (D) slake (E) revise

A student who didn't know the meaning of slake, the right answer, could still do well in college. Even if he didn't know the meaning but knew the precise meaning of *whet*, he could get the right answer by process of elimination. On the accompanying answer sheet, E.T.S. informs us that of students getting a verbal score above the median of 450 on this test, only 21 percent settled for *slake*.

In an analogies section, in which the student is asked to "select the lettered pair that best expresses a relationship similar to that expressed in the original pair," come these pairs:

SWILL:SWINE—(A) roe : fish
(B) coop : poultry

(C) mutton : sheep
(D) pesticide : vermin
(E) fodder : cattle

If a student had read rural English novels or grown up on a farm, he would surely know what *swill* meant and make an instant connection with *fodder*. Only 28 percent of the near-median scorers got E, the right answer. Since 450 is not a high S.A.T. score, perhaps some poor farm boys succeeded where urban or suburban students failed.

To give the devil his due, it must be said that the S.A.T. asks for an ability to make fine distinctions, as in this analogy:

IMPREGNABLE : AGGRESSION—
(A) imperfect : revision
(B) invincible : defense
(C) inequitable : criticism
(D) indivisible : separation
(E) immutable : preservation

Thirty percent of the near-median scorers chose the right answer, D. Just as something is *impregnable* against *agression*, so is something *indivisible* if *separation* is tried.

S.A.T. sentences, with their blanks filled in correctly, read well enough in this one released test; even the reading passages do not assault the ear of the sensitive reader. But even though verbal sections of the S.A.T. are well enough written, the test is still obnoxious. Its verbal part, like other objective tests in language, does not ask for writing; and because it doesn't, the act of writing has withered in our schools.

Reading questions on the sample S.A.T. try, with considerable success, to trick the student. Here is a paragraph from a longer passage on the sample test:

The way of the desert and the way of the jungle represent opposite methods of reaching stability at two extremes of density. In the jungle there is plenty of everything life needs except mere space. Everything is on top of everything else; there is no cranny which is not both occupied and disputed. At every moment, war to the death rages fiercely. The place left vacant by any creature that dies is seized almost instandly by another, and life seems to suffer from nothing except too favorable an environment. In the desert, on the other hand, it is the environment itself which serves as the limiting factor. To some extent the struggle of creature against creature is mitigated, although it is of course not abolished even in the vegetable kingdom. For the plant which in the one place would be strangled to death by its neighbors dies a thirsty seedling in the desert because that same neighbor has drawn the scant moisture from its spot of earth.

The question pertaining to this paragraph is:

Which of the following is (are) true of both the way of the jungle and the way of the desert?

I. They are characterized primarily by the struggle of creature against creature.
II. They are reactions to hostile environments.
III. They result in population control.

(A) II only
(B) III only
(C) I and II only
(D) II and III only
(E) I, II and III

What is a student, in a limited time, to do with these choices? The instructions tell the student to answer "on the basis of what is *stated* or *implied*," but there is both a statement and an implication to reckon with. The test makers undoubtedly intended to lure many students to "I. They are characterized primarily by the struggle of creature against creature." For not only is the struggle a clear characteristic of both desert and jungle, it is also emphasized in the last sentence. But strong as that temptation may be, the student who yields to it will be wrong. For the test makers have cleverly inserted the word *primarily* and linked I with the unacceptable possibility of II. To score well, the student must have his guard ready and high. The right answer is *(B) III only:* "They result in population control." Perhaps the right answer is got by process of elimination. Or perhaps it is obtained by finding the faint implication in the phrases "reaching stability" and "limiting factor." To get the answer in a short time does indicate something akin to the skill of a detective. But it does not indicate that a student could write a searching term paper. Getting the right answer does not measure the depth of a student's mind. Of those who scored 450, 20 percent succeeded.

Vainly, E.T.S. tries to measure writing ability objectively in a 30-minute part of the S.A.T. called the Test of Standard Written English (T.S.W.E.). The test asks the student to spot the grammatical errors in underlined sections of sentences; it also asks him to choose from among five possibilities the best way to phrase an underlined part. A student who can spot an error in an underlined section may still commit the same error in the frenzy of unaccustomed composition. Even if a student has a sure knowledge of grammer as shown on an objective test, the test still won't determine whether he can write an essay. For writing requires not only grammar but ideas and the ability to organize material. These abilities show up only in actual writing. The test is in many ways an exercise in futility. The student's score is not counted in the S.A.T. but sent to the college, supposedly to determine what level of freshman English the student should take. Good colleges will not use the score for placement purposes, as E.T.S. recommends. Most colleges administer their own written essay exams to determine what level of composition course a student needs.

E.T.S. does ask for a 250-word essay in one of the five annual offerings of the optional Achievement Tests in English Composition. But of the 1.4 million students who took the S.A.T. in 1977–78, only 85,000 students wrote the essay. In the multiple-choice English Achievement Test, the student marks errors in "diction," "usage," "idiom," "wordiness," "sentence structure" and "metaphor" without ever demonstrating that he can avoid making such errors in his own writing. Many good writers, students or not, may be unsure of academic terminology but write well out of their sense of language. That kind of talent would do poorly on the test.

The test encourages schools to teach primarily by terminology, since the final test asks for it. For example, the student must choose from several possibilities the right rephrasing of a sentence if a subordinate clause is changed to a participle, or if one sentence is changed to a clause. He is told, in effect, that writing is a juggling act, without purpose.

The testing services that dominate our educational system may be able to justify tests in mathematics, but they shouldn't be allowed to sap the strength from our language with objective tests in writing. If we are given another generation of tests, writing will become a rare art. The responsibility for salvaging writing falls to the American university, which instituted the objective tests for college admission 25 years ago and sanctioned their use throughout the school system.

The antidote is to restore writing requirements on college entrance exams. But the task is formidable: E.T.S. has a vested interest in protecting its contracts; it is marshaling every "social science" argument it can in defense of existing tests and challenging the "scoring reliability" of essay tests.[2] Justifying the multiple-choice approach, E.T.S. declares that "students who recognize the problems in the writing of others are likely not to have those problems in their essays, an assumption confirmed by careful research." English teachers know this is often not the case. The real problem seems to be money. E.T.S. throws up its hands at the cost of grading written exams, but surely a way exists to pay readers to read written essays.

No critic has defined the shortcomings of the S.A.T. better than Ralph Nader. "E.T.S. has us all locked into a test that doesn't look for creativity, stamina, motivation or ethics—which are the four qualities on which man's greatest achievements are based." Our mania for measurement, our naive assumption that we will discover a true meritocracy through objective standards, contradicts the original American idea. When Jefferson wrote of a "natural aristocracy" emerging from an elementary and then a "higher degree" of education in preparation for the university, he thought that "worth and genius would thus have been sought from every condition of life." "Virtue and talent" were to be relentlessly encouraged and sought in schools at all levels. What worth, what talent, what virtue, does objective testing measure? Rather than rewarding visible intellectual work done in the classroom, the tests reward a cleverness that even the clever can doubt.

Though E.T.S. announces that the S.A.T. is not designed to judge the "worth" of anyone, the tests set implicit standards of worth by becoming a passport to education, income and social status. An S.A.T. score—the score on a single test—can set the direction of a lifetime. Though there is increasing skepticism about the S.A.T., the score can still weigh heavily, E.T.S. admits that a 60-point difference in verbal scores is statistically meaningless.[3] Yet a 60-point difference between two students can mean that one will get into a particular college while the other will be denied admission. To "learn" to take the test, students spend valuable hours in cram schools and crash courses—and even attend summer school. Though the coaching sessions can be helpful, they also teach an appropriate cynicism toward "the system." Substantial education—the history of man or nation—is set aside in order to beat the test. The test is "the system," to be overcome by studying vocabulary lists, by taking practice tests, by learning the technique of test taking. Resentment and cynicism are two lessons taught by the American way of testing.

---

[2] Could grades given by individual readers be any less reliable than scores given by E.T.S.? Hardly. E.T.S. admits to a "standard error" of 32 to 50 points on its S.A.T. scores.

[3] The 60-point difference is meaningless because 60 points could have been attained by such external factors as guesswork.

[From the N.Y. Times, September 9, 1979]

Test-Coaching: Is It Worth It?

(By Nancy Rubin)

Frank Carmona sat intently in a makeshift classroom at Manhattan's Barbizon Plaza Hotel as a Columbia Test Preparation Institute teacher, Steven Stumpf, outlined strategy for taking the business section of the Graduate Management Admission Test. "Don't think any more than you have to," advised Mr. Stumpf, "The best approach is to eliminate the bad answers first, then concentrate on the remaining choices."

Several blocks south, Charles Barrett emerged from a tape drill session at the Stanley H. Kaplan Educational Center. "My whole life depends upon passing this test," said Mr. Barrett, an American studying medicine at the University of Noreste, Mexico, who is preparing for the first part of of the National Medical Boards. "I'm hoping to pass the test and transfer back to an American school, and Kaplan's course has been a godsend."

Test-preparation schools are experiencing unprecedented growth among students of many age levels with a wide variety of special interests, despite attempts by some educational groups, Federal agencies and the standardized-test industry to destroy their credibility. While the twin controversies of the testing issue—the rationale for standardized examinations and the belief that those examinations should be publicly accountable—have divided the educational community and led to demands for a Federal "truth-in-testing" law, many students have embraced the concept of standardized-test preparation as a necessary, even rightful, addendum to the educational system.

The issue flared into national attention this summer when the New York State Legislature passed a "truth-in-testing" measure that calls for the public release of test questions and answers as well as an explanation of how those answers are scored. The measure was enacted in the hope that it would "equalize" the testing experience for all students—NAACP had argued that "many students cannot afford these expensive coaching courses and are therefore denied equal access and opportunity"—and provide them with better guides on their academic strengths and weaknesses.

In the last decade, test preparation has developed into a $60 million industry, resulting in a proliferation of test booklets, courses and institutions.

The nation's largest test-preparation school, which is owned by Stanley H. Kaplan and has branches in 88 cities, experienced an annual sales growth of more than 25 percent in one year, from $7 million in 1977 to $9 million 1978. Smaller schools, like the John Sexton Test Preparation Center, which served about 6,500 students and had sales of $650,000 last year, and the Amity Testing Institute, which offers test preparation in 40 cities and grossed more than $550,000 last year, have also begun to offer new courses and expand their locations. At the same time, many corporate, private and nonprofit educational institutions have expanded test-preparation activities.

Citibank, for instance, recently engaged the Sexton School to teach courses in the Graduate Management Admissions Test (GMAT) and the Law School Admissions Test (LSAT) to its employees. "We perceived test preparation as a need and are hoping it will enable our employees to do better on standardized tests," said Eileen Sini, program manager for the Foundation Center for Adult Studies at Citibank.

Formal efforts to evaluate the test-coaching industry have yielded few clear-cut answers. Several studies conducted by researchers at the Educational Testing Service (ETS), the nonprofit testing corporation that administers standardized college- and graduate-school admissions tests to more than two million students annually, have found little evidence that test-preparation courses raised student scores by more than 10 points, a minimal gain, the testmakers say, compared against the cost of a course, some $200 to $300.

A 1972 survey conducted in 12 high schools by two ETS psychometricians, Lewis Pike and Franklin Evans, found that a 21-hour coaching program did have a positive effect on test scores for students on certain parts of the mathematics section of the SAT.

A 1978 review by Mr. Pike also concluded that short-term instruction produced meaningful improvement in test scores, for the mathematics part of the SAT, although the effect of short-term instruction on verbal SAT scores was still unclear.

This spring the Federal Trade Commission released a report examining the effects of coaching schools upon standardized-test scores. It found that "students who are coached for the SAT have, on average, higher grades than students who are not coached." The report which used previous data gathered by the Boston office of the F.T.C. to examine the validity of coaching, school claims indicated that coaching was especially helpful for "underachievers," often raising their scores by 25 points on both the math and verbal parts of the SAT. Underachievers were defined in that report as those with high grade-point averages but low standardized-test scores.

However, the final report was disclaimed by the F.T.C's Bureau of Consumer Protection as being statistically invalid and failing to control for differences in personal and demographic characteristics of the students.

According to Mr. Pike, who is no longer employed by ETS and was recently hired by the National Institute of Education to conduct educational testing research, the initial findings of the F.T.C. study are not surprising, because he believes that students often do not do their best because of the format of the test itself. He says that test anxiety, unfamiliarity with the question format and the artificiality of the question types themselves are powerful obstacles to valid test results. For this reason, he asserts, test preparation can indeed prove useful. "I don't think a student can artificially raise his test scores above his ability through test preparation," Mr. Pike said, "but he might bring his scores up to where his ability really is."

It is precisely in this area, where students exhibit a marked discrepancy between their academic performance and their test scores, that the test-preparation schools seem effective. According to Ellen Kurtz of the Livingston College Board Review, in Livingston, N.J., for instance, students who have high grade-point averages but are scoring between 450 and 600 on each section of the test are more likely to be helped by test preparation while students who have scores above 600 seem to show less dramatic improvement.

"A very bright student with scores already over 600 may not be someone we can help because he's already picking up intuitively what to do from the types of questions being asked," Mrs. Kurtz said, "but our course really is more effective for the average student who is more apt to be thrown by the nature of the test itself."

Other test-preparation educators maintain that their courses are particularly helpful when students are apprehensive, or tend to "freeze" on standardized tests. "By increasing the students' academic knowledge, showing them examples of old question types found on previous tests, their fears are often greatly reduced and they just naturally perform better on the next test," explained R. C. Matthews of the Matthews School in Manhattan.

In recognition of these subtleties, some schools, such as the Kaplan institutions, while seeking to improve student test scores, discourage students with poor grade-point averages from applying to prestige schools, reasoning that their past academic performance will impede their acceptance. At the same time, most test-preparation schools are careful to make few claims about score gains, although they often say privately that increases of 50 to 100 points are not unusual.

According to the College Board, which employs the Educational Testing Service to design and administer the SAT, LSAT, the Graduate Record Examination (GRE) and other standardized tests, the scores are not nearly so important a factor in school admissions as most students believe. "I know of no school that selects or rejects students on the basis of test scores alone," said Fred Hargadon, chairman of the board of trustees of the College Board and dean of admissions at Stanford.

On the other hand, Allan Nairn of the Public Interest Research Group, a Ralph Nader organization that is investigating the E.T.S., said: "The fact is that the tests are important and are used in a variety of ways by schools to determine entry. And at least 20 percent of the colleges still use them as automatic cutoff points."

Whether colleges are using standardized tests as a primary gauge for admissions or not, students evidently believe in their significance, and it is to this need that the coaching schools address themselves so successfully.

The schools vary widely in the technique and approaches to standardized examinations, but virtually all indicate that "testmanship," or specific test-taking skills, are important curriculum components. "Schools often don't teach students how to take tests because they have so much basic subject matter to get across," said Joel Piagenz, co-director for the New Haven-based Amity Testing Institute.

Many schools incorporate test-taking techniques within their teaching of the basic subject matter, stressing such skills as when to guess, how to recognize verbal mistakes, how to use mathematical short-cuts, and when to implement precision reading. At the Columbia Test Preparation Institute, for instance, students are drilled in the analogy questions commonly found on standardized tests. Within the verbal portion of their classroom instruction, students are briefed on the "traps" that analogy questions often contain, such as the "reverse meaning" choice, the "word association" choice, and the "ungrammatical answer" choice. Having learned to recognize and eliminate those inappropriate choices, the student is in a better position to determine the correct answer.

One of the most valuable aspects of the test-coaching schools, their advocates say, is the way they reduce anxiety. "After all the rumors I've heard about the difficulty of the test, this course has made me feel much better," said Lynn Goss, a student at the Columbia Test Preparation Institute, who was preparing for the G.M.A.T. examination. "The instruction was a logical presentation of the material covered on the test, and once it provided me with explanations about how to find the answers, I was no longer nervous."

"A time-pressure test seems to rattle many students and destroys their ability to do their best," said the coordinator of the L.S.A.T. test-preparation center at Rutgers University, Janet Finklestein. In the Rutgers program, as in many other programs, the instructors try to overcome that anxiety by offering the student intensive work on the subject matter as well as expressions of encouragement.

To dispel student anxiety, some schools, such as the Sexton Center, provide remedial mathematics courses in addition to the standard program. Others, such as the Dulac School in Manhattan and the Guidance Center in Santa Monica, Calif., offer individualized counseling and psychological reassurance. At the Guidance Center, for instance, the director, Anne Salzman, attempts to reduce worries by counseling her clients and discussing their anxieties with parents. In addition she often diagnoses her students' academic weaknesses and outlines a program of study for them in needed areas.

Most test-preparation schools also offer students a strenuous review of basic academic subject matter. At the Test Prep Inc. school in Chicago, for instance, students have 34 hours of classes to review basic mathematical and verbal skills.

At the Kaplan schools, where students get up to 10 hours of live instruction but are encouraged to participate in 200 hours of taped drill exercises, a wide variety of practice material is presented.

Mr. Kaplan maintains that his students not only do better on standardized examinations but also increase their basic skills permanently. "I haven't just improved their test scores, I've made them better students," said Mr. Kaplan, who asserts that his courses are effective because they teach students basic concepts over an extended period of time.

The cost for test preparation varies widely. As a rule, SAT preparation is less expensive than other types of graduate preparation, ranging from $80 at small schools like the Livingston College Board Review to $275 at the Kaplan schools. Fees for professional license examination courses, such as the Education Committee for Foreign Medical Graduates test, which clears foreign doctors for practice in the United States, come to $800 at the Kaplan schools.

TIPS FROM THE COACHING SCHOOLS

1. Plan your time carefully. Spending too much time on a question can be as detrimental as a wrong answer.
2. Answer the questions you know first. Then, If you have time left, return to the unanswered questions.
3. There is only one right answer. If you identify it, do not waste time working through the other possibilities. Go on to the next question.
4. Guess shrewdly on answers you do not know. Wrong answers carry a larger penalty than unanswered ones, but your chances for a higher score will be improved if you guess—provided you can first eliminate at least one choice.
5. In reading-comprehension passages, underline important information as you read through the test. There is no penalty for marking the worksheets, and it will help you recall important information. Likewise, in mathematical sections use the white space as a worksheet.
6. Pay particular attention to words like "but," "not," "however," and "therefore." They are key words that often signal major thoughts of a passage.

688

7. Use mathematical short-cuts—cancellation of fractions, estimation, removal of decimal points and so forth—wherever possible.

8. Memorize math formulas. Formulas such as A = l × w for the area of a rectangle are often given at the beginning of the test section, but it is quicker to have them in your head.

9. Don't think any more than you have to. Don't be sidetracked by secondary answers or answers that are true but not directly related to the central question.

10. Get a good night's sleep and bring candy and a sweater to the test. The examination is tiring and such means can help you conserve energy.

Because of the cost of test-coaching schools and the widely held belief that coaching works, the Department of Health, Education and Welfare has allotted funds to the Health Resources Administration Office of Health Resources Opportunity to enable disadvantaged students interested in the health professions to take extra courses and test preparation. This year the department awarded more than $3 million to disadvantaged students for coaching at 21 universities through the Health Resources Administration. It was the largest amount of money ever granted to those programs.

In one program, at the University of Texas Health Science Center in San Antonio, 50 minority students with high academic averages studied and received preparation for the Medical College Admission Test and the Dental Admission Test through the Kaplan schools.

While some advocates of the "truth-in-testing" legislation pending in Congress maintain that it would make the testing industry more accountable and would allow the answers and techniques used on standardized tests to become more accessible to all students, not just those able to afford a preparation course or commercial preparation materials, directors of test-preparation schools have expressed little concern about the bill.

"Whether they impose a Federal "truth-in-testing" bill or even abandon standardized tests altogether will make little impact upon us," said Emanuel Federbush, director of the Columbia Test Preparation Institute. "We may have to make adjustments, but no matter how schools decide to judge their applicants, students will always be looking for a little extra help."

TESTIMONY OF DR. LEWIS W. PIKE BEFORE THE NEW YORK STATE SENATE AND ASSEMBLY HIGHER EDUCATION COMMITTEES, MAY 9, 1979, ALBANY, N.Y.

I am Dr. Lewis Pike, an educational researcher currently employed as a Senior Associate at the National Institute of Education. I wish to thank you for the opportunity to express my views regarding proposed legislation on standardized testing. The testimony I am presenting is in a private capacity. No official support or endorsement by the National Institute of Education is intended or should be inferred.

My professional background for providing this testimony includes the following. I received a Ph.D. in educational psychology from Teachers College, Columbia University, in a program that emphasized educational testing and measurement. My professional work experience includes 14 years at Educational Testing Service, two in test development and twelve in research. While in test development, I was in charge of developing all verbal aptitude test materials for the Scholastic Aptitude Test, the Graduate Record Examination, and several other tests. My work in the research division was directed primarily to various properties of such tests as the SAT and the GRE, usually involving such questions as test fairness, test validity, and possible coachability of tests.

I have long been, and continue to be, an advocate for the use of standardized tests for appropriate purposes. In particular, I believe they can provide an important service both to colleges and to students seeking admission to them. However, the very importance of such testing suggests to me that no reasonable effort should be spared to ensure that both constituencies, colleges and students, are well served.

My comments on the draft legislation under consideration will be addressed specifically to Section 342, "Disclosure of Test Contents." Essentially, this section would require that test copy be made available to the public within 30 days of the release of test scores, and that students be provided, upon request, their answer sheets.

I strongly favor this provision for "open testing", as opposed to "secure testing", for three reasons:

> *First:* The assumption that students and others wanting to see test copy (after test administration) must demonstrate "the need to know" is inappropriate, because it puts the burden of proof on those making the case for openness. In

our society, the burden of proof more properly rests with those advocating secrecy, particularly where a critical public trust, such as college admissions testing, is involved.

Second: Arguments that I have seen thus far for maintaining blanket test security after test administration are not convincing. Of six specific objections to open testing raised recently by Educational Testing Service, four boiled down to a matter of added cost, and the other two (test-equating and item-pretesting needs) are readily met by exempting items in those categories from being made public.

Third: Given the importance of standardized tests in college admissions decisions, it is essential that the tests be as open as possible to public scrutiny. For such critical review there is no effective substitute for access to test copy and associated test data.

Let me briefly expand upon those three points. In doing so, I shall refer to two documents. One is a paper by John Fremer and Alice Irby, of Educational Testing Service, entitled, "Why Should Tests be Secure." It was distributed just last week at a conference of the National Consortium on Testing, at a panel discussion of "Open versus Secure Testing". The other is a College Board publication, "Report of the Commission on Tests: I. Righting the Balance," issued in 1970.

The first point: The burden of proof should rest with those advocating the maintenance of test security.

Noting ETS's multiple constituencies, the question is phrased by Fremer and Irby as follows: "How can test publishers meet what may appear to be conflicting needs of institutions and associations on the one hand, and the public's need to know?" Now, the concept of "the need to know" is more typically heard in reference to such items as classified military information, and its use assumes an understanding that the burden of proof for gaining access to the information rests on those wishing to see it. I would maintain that when the public's access to test information is concerned, and scores of great personal significance to the tested individuals are involved, the opposite assumption must prevail. A provision for "open testing", such as that given in the proposed legislation, is consistent with shifting the burden of proof to those claiming that extended test security is necessary.

I feel this is particularly true for an often-overlooked constituency, the students required to take standardized tests in order to gain admission to college. A concern for this constituency has been strongly expressed to the College Board by a Board-sponsored Commission on Tests in Volume I of their report, Righting the Balance. This report included the following comments:

> As the Commission's work proceeded, it became apparent that, in its members' opinion, the Board's traditional services reflected primarily and purposefully the interests of the Board's member colleges; and that, while these were not necessarily inconsistent with the interests of students, the latter, being served secondarily and incidentally, were being served less well. (Page XV.)
>
> While in reality no can solve anyone else's problems of choice, the College Board can give the students faced with the difficult decisions surrounding the transition out of high school support equal to that which colleges receive. In short, the Commission thinks that a symmetry or balance should obtain between the services that the Board offers to potential entrants and those that it offers to colleges. Just as individual schools and individual colleges rightfully regard their students as clients, so should the member schools and colleges of the College Board have the Board regard all potential entrants in programs offering opportunities for postsecondary education as a clientele whose interests and needs are to be served and met as fully as are those of the Board's institutional clientele. (Page 57.)

In making specific recommendations, the Commission on Tests suggested that students should be provided much more information that would allow them to evaluate prospective colleges. The present legislation calls for a natural extension of this concept, in which the students would also be allowed a more adequate basis for evaluating the admissions tests that so directly influence their lives.

The second point: Present arguments for maintaining blanket test security long after test administration are not convincing.

Fremer and Irby listed six reasons why test security practices should be continued. One was "Reuse of Test Questions," with cost noted as a major objection to limiting the use of test questions to a single round of testing. Three of the other reasons, given as "Fairness," "Meaningfulness of Test Result," and "Discourage Teaching to the Test" are all based on the premise that test items would have to be reused, after having revealed to the public. But that premise holds only if it is indeed demonstrated that using all new questions for tests now treated as secure would be prohibitively expensive. Thus all four reasons may be reduced to a single consideration, cost. With the many cost factors other than the development of test items that must be covered by test fees, and the likelihood that a substantial part of most secure tests already consists of new items, I would expect that the claim of excessive cost for developing all new items would be very difficult to support.

The other two reasons given for insisting on test security are "Continuity of Test Scores" and "Presenting of new Test Questions." As I noted earlier, these objections are readily met by exempting items in those categories from being made public.

The third point: It is essential that tests be as open as possible to public scrutiny; for this, there is no substitute for direct access to test copy.

Two substitutes for direct access to test copy and data have been suggested. One is the use of external test committees to ensure the quality and appropriateness of resulting tests. While this is a commendable practice, it is primarily a process consideration. The product of interest is the final test itself. Further, such committees do not directly represent the student as consumer. More strongly emphasized as a substitute is the availability of extensive descriptive information about secure tests. This, too, is good to have, but again it simply cannot replace seeing the test itself. As a researcher, I have consistently found that my work with tests such as the SAT, GRE, and the Test of English as a Foreign Language could be fruitful only after I had spent considerable time reviewing and reacting to direct test copy, interacting with individual test items, relating these to one another and to test directions, and so on. If I were a consumer (test-taker) or a legislator charged with representing consumers, I would not willingly settle for less.

I shall conclude by citing another of the Commission on Tests' statements, which reads: * * * the tests and associated services intervening between high school and college, especially those of the College Board that touch directly as they do the lives of over a quarter of the nation's young people, are a matter of some importance, not just to the Board, those young people, and the institutions they attend, but also to society at large. So large and pervasive an enterprise should, as should the educational system it supports, be managed in the public interest. The challenge for the College Board is to discern that interest and effectively serve it. (Righting the Balance, Page 32.)

This charge, of course, could apply to the testing enterprise at large. It is my belief that the public interest referred to in the statement can be best served if open testing practices are followed, allowing the public direct access to test content. I believe the legislation before you, particularly Section 342, is an extremely important step in bringing about such open testing.

REFERENCES

Fremer, John, & Irby, Alice J. Why should some tests be secure. Preliminary paper presented at the Spring Conference of the National Consortium on Testing, Arlington, Virginia, May 1979.

Commission on Tests. Report of the Commission on Tests: I. Righting the Balance. College Entrance Examination Board, Princeton, N.J. 1970.

Mr. **Weiss**. I think that Dr. Berendzen has in essence raised the very clearcut argument of those who are opposed to H.R. 4949, and I would like to see if we could in the course of the time that is left to us discuss or address ourselves to some of the questions that he has raised and perhaps we have not specifically focused in on.

It has been suggested by many, it has been suggested by others in the course of the hearings that we have held up to this point, indeed it has been suggested by an editorial in the Washington Post by some others, that the standardized tests have in fact been responsible for bringing into higher education minorities and others who up to that point had been excluded, and therefore the suggestion has been made that when we seek to open the process by disclosure, because that is all that H.R. 4949 does, it is a disclosure.

The piece of legislation mandates nothing beyond disclosure, that what we are doing then, at least the suggestion is, is working against the efforts of minorities and others who had found it difficult to gain entrance into institutions of higher learning previously.

I would like to see if we could get some response to both parts of that, (a) whether in fact it is the standardized test, standardized tests which have been responsible for bringing minorities into the higher education system, into the colleges and universities and, second, whether the adoption of H.R. 4949 would have the effect of now turning the clock back and excluding those people.

Do you have any comments on that, Dr. DeWinter?

Dr. DeWinter. Yes, sir. I really do not see how these tests would have all by themselves brought about the change in the higher education in bringing additional minority students to the colleges.

One thing that concerns me very much about the tests is the extent of our knowledge of them and the relationship between the knowledge of the tests and what they do, whether it be for any student from any area, and then how that affects our judgment of the decision of the admissions office.

It seems to me that disclosure, as proposed in H.R. 4949, would simply increase our own understanding of it, but it would still not substitute for the kind of judgment that we would have to exercise in interpreting the individual candidates.

So I really do not believe that it would act in a way to exclude any individuals, and particularly minority students, from higher education.

Mr. Weiss. Would some of the others like to comment on that? Dr. Franklin?

Dr. Franklin. Yes.

In response to your question about test scores improving the admission of minorities in schools, I find that a little hard to believe because we already have testimony that minorities have scored one standard deviation below and there has not been any admissions of students because of these test scores alone.

It is because the test scores are low and because there are other factors involved that minorities have been admitted to programs and just a general awareness of the inequality in education.

So that test scores have not been the savior or deliverer of minorities into the arms of higher education. There have been a host of other political and educational issues that have delivered minorities into higher education. So test scores certainly are not the messiah in that regard.

Second, the disclosure of test questions certainly raise psychometric issues, but it is interesting that all we have heard is what harm it is going to do. We never hear that the industry is going to mobilize to try to handle the particular issue of what disclosure may do, to develop the necessary techniques and necessary tests.

It seems to me that disclosure will allow students to learn their strengths and weaknesses. It should contribute to quality education. It should contribute to the development of better curriculum.

It should allow us to know as independent researchers and independent educators from the testing agency exactly what areas of knowledge people, students, are not being sufficiently trained in and what processes are necessary to help students to learn better. It should improve the whole pedagogic process.

It is a shame that very often the tests are represented as a messiah, as a deliverer, as a panacea. They are not. I think what we are trying to do is to improve testing to the point where it becomes an integral part of quality education.

Mr. WEISS. Dr. Loewen?

Dr. LOEWEN. I think we also have to recognize that there is already a great deal of knowledge about tests available in society. The student who, let's say, locally goes to Walt Whitman High School has all kinds of knowledge about the tests offered him or her as part of their curriculum in the school. The same is true of the person who goes to Phillips Exeter.

There are coaching schools which ETS is now retreating on and is aware that they improve scores. Where is the knowledge about how tests are, their form, how to take them, where is this knowledge located? It is located in the upper middle class in urbane and urban areas.

The function of H.R. 4949 is to increase the knowledge about the tests available to all Americans. I cannot see how that would decrease the opportunity of rural or regionally dispossessed Americans, if you will, to take the test.

I think it will only amplify the opportunity that the tests purport to make available.

Mr. WEISS. Ms. Simmons?

Ms. SIMMONS. I believe, based on my contacts with students across the country, that as a matter of fact the standardized tests have not served to bring more people into institutions of higher education. It has served to keep a lot of persons out, also out of professional schools.

One of the things we must also look at, as I mentioned in my testimony, that the bill appears to be directed toward postsecondary. However, there has to be some attention to elementary, secondary, in order to get the youngsters up to where they can go into postsecondary.

One of the things we have found out in segregated schools that we do still have across this country is that in the schools that are in the upper middle income white areas, that a number of times youngsters get an opportunity to take standardized tests early on so that by the time they reach 11th or 12th grade they have a facility for taking those tests.

That is not true in the Bedford-Stuyvesant and Harlem and other areas that are so-called ghettoized areas. We do believe that standardized tests are used to exclude. They impact disproportionately on low-income persons regardless of whether they are black, white, or brown. The overemphasis that is being placed on standardized tests does impact disproportionately.

May I make a comment about some of this in my testimony? We have had occasion a few years ago to look at a school system in Virginia where our branch had collected for me about 75 report cards. On those report cards they had passing scores of students and yet these students were partially promoted.

When we went to the school system, there was a public forum which we had called for. We showed report cards, the local paper showed copies of report cards of a person who had passing scores on a report card.

The school system said that the standardized test was not the sole criterion. However, they could not respond when they were shown report cards of persons who made B's and C's on report cards and who were not promoted.

So we are consciously aware of the fact that even though we were told that the score itself was not the determining factor, it was, and we had just a fistful of report cards to validate that.

Dr. **Newsom**. Mr. Chairman and members of the committee, this is a very difficult situation for me, one as a minority and as an editor, and particularly as one who has some understanding of the cultural background from which these tests are developed.

These tests are not without bias. There is no test made without bias. So these tests are obviously biased in favor of the majority group, in favor of the high socioeconomic group, in favor of those who live in a certain part of this Nation. Therefore, minorities for the most part suffer and are excluded.

My concern is that there must be some development more and more working toward reducing the bias that appears in these tests.

What we have said here, that we should modify in some respects, we should augment, expand others, and change these in order that they will be more meaningful and do what I think the Congress wants to do. We cannot do it the way it is presented at this time.

Mr. **Weiss**. But, Dr. Newsom, the hope is and the expectation is that if the mystery is taken out of the tests, if the tests become a matter of open and public knowledge, if the student himself and herself has access not just to the questions but also to the answers on request, if researchers, sociologists have a chance to really examine the tests and what went into the process of making them up and evaluate some of the in-house studies that are made and undertaken by some of the testing services, that just that process of opening it up, disclosing it, will in fact allow sufficient critiquing, information if you will, from outside sources to be borne in upon those who provide the tests now and who make them up and compose them so that in fact the tests will more accurately test the abilities and the capacities of students across the country rather than particular regions.

That is why the test really focuses on openness and disclosure rather than mandating any specific action.

So I am a little perplexed as to why you think there should be greater study of that process unless what you are suggesting really is that the legislation ought to do more than what it does and we have deliberately held off from doing more because we think that would subject us perhaps to valid criticism that we were trying to interfere with the educational process in some fashion.

Dr. **Newsom**. Well, basically now you are, but I think this is good. I think what we are asking is that it be studied more by the social scientists with the Congress in order to provide a technique, a law or a rule or regulation that will do what you say you are trying to do.

We don't believe this will do that. This will continue to do what it has done. That is our belief.

Ms. **Simmons**. Mr. Chairman, I certainly cannot endorse or call for an additional study. I would have wished the bill would have gone further. I know it is absolutely mandatory that we have truth in testing legislation with disclosure.

I want to say for the record that I hope in the future that we will have a bill before the Congress that will establish a Federal regulatory agency for the testing industry similar to the Federal Communications Commission, Food and Drug Administration, Environmental Protection Agency, for the protection of consumers of tests, and evaluative assessment instruments.

Because more and more, if you will recall, the Congress is calling for more assessment, particularly in the elementary and secondary education.

I believe that with a call for more and more assessment, there is also responsibilities to regulate.

Mr. WEISS. Did you receive in the course of the hearings, especially from the testing companies, the assurances that any social scientist who wants to do research in the area of testing has all the materials that they, the testing companies, have available for their use?

We have had indications from some of our witnesses who attempted to receive or gain access to some of those materials that at the least that is an overstatement on the part of the test industry.

I wonder if you have had any experience in attempting to receive materials, access to materials, and what that experience is?

Dr. LOEWEN. Yes. I have had particularly disappointing experience in dealing with the Educational Testing Service and the College Entrance Examination Board that oversees some aspects of ETS.

Last fall I was attempting to do research on testing. I contacted people at ETS about the potential availability of what they call ethnic data by which they mean data by race, perhaps Hispanic-Americans included as against Anglo-Americans and black Americans, and also data available by State.

I was referred by ETS to George Hanford who was then senior vice president and now I understand is acting president of the college board.

I wrote Dr. Hanford on October 23, 1978, and asked him about that availability. I explained to him that I was a social scientist and what the Center for National Policy Review was and that I was coming from a proaffirmative action background.

His reply November 10, 1978, was:

> Data relating to the PSAT and SAT do come under college board authority and are subject to a policy administered by me which does not permit the regular aggregation and reporting of data regarding the test performance of minority students.

So he claimed, in other words, that they do not even aggregate data on the basis of minorities and they do not permit the reporting of it.

It was then claimed to me just this last week that that policy had changed. So last Friday, 3 days ago, I phoned the college board to see if it in fact had changed. I was assured by Darryl Greer, the policy planning officer at the office of the president, that the policy had not changed.

He said we were in a process, then, of reviewing that policy and the review is still ongoing.

He said, "The board has asked for years whether it should aggregate such data, and if so, what its responsibility should be to any differences we might find."

In other words, he is claiming that the board doesn't even know that there are differences. How then can ETS or the college board assure us that there is no test bias if they claim they have not aggregated data and they have not set up what the response should be to any ratings or other differences that they might find?

But in my case the answer to your question is that as of last Friday it is impossible to get this kind of data, at least for me, and I don't think there is anything wrong with my credentials. They explained this to me as their policy.

Ms. **Simmons**. May I respond to that, too.

I talked with five psychologists across the country when I formed the national task force. Dr. Robin Williams, who is a psychometrician felt there was a need that section 42(b) in H.R. 4949 was probably the best section in the bill. He felt there was a need for a mandate for the commissioner to get relevant data over the period of time to see what was going on.

He said he was reliably informed that the data was available in the computers, but there is a feeling among the professionals in trying to get this information from the companies that they will not release the information unless there is a congressional mandate, and that if the commissioner had information on what was happening through the use of the test, the commissioner would then be in a position to report to the Congress and to make a determination if further action was needed.

So that in talking with various persons who have credentials, who need access to this information, they have not been able to obtain it.

Mr. **Weiss**. Thank you.

Dr. DeWinter, I am sure that you have had occasion in the course of your work, in discussing this particular area that is your area of specialization, and that is admissions generally and standardized testing specifically, with others in the field, other admissions officers.

Generally, can you describe as to where we seem to be in general attitudes? Does there seem to be a stronger commitment today on the part of admissions officers to the standardized test than there was 10 years ago or a lessening? What is happening? Is there any kind of fluctuation that is going on?

Dr. **DeWinter**. Congressman Weiss, I don't really see any significant fluctuations in the importance given in the standardized tests. It seems to me that admissions officers throughout the country, I would say with whom I have been associated, really do try to take all of these factors into consideration, including the SAT's.

So in that respect I don't find that there is an increase. I think what is happening, though, and this is why I am endorsing the legislation, is that in the public's mind they definitely take on a tremendous amount of significance. There is almost no way that you can really deal entirely with the question in explaining a particular decision to an applicant. It always comes back to those scores.

So in that respect they really have a tremendous impact, at least as perceived by the public. I believe that with an increased understanding of the scores, that is through the disclosure, that actually we can lessen that impact on the public.

I think, also, it would make us be very careful so that no abuses are made of these tests.

Mr. **Weiss**. You have come to your judgment, obviously, quite independently and you are not speaking on behalf of your university or anyone else but yourself?

Dr. **DeWinter**. If I may add something, on the question of validity and the use of it, we have talked a great deal about tests in relation to the predictive ability of the students.

I think there has to be a healthy caution about that also because it is just not very clear to what extent those scores really have been valid in predicting the success of a student.

Just before I came down I asked a question at my college. I said: What would be the scores, for example, of the top 10 percent of the students?

It was quite clear that there were variations in all of these scores. There were as many, say, among our Phi Beta Kappas, there were as many who were scoring between 500 and 600 as there were between 700 and 800.

So I think we have to be very careful with the predictive validity, saying that the higher the scores, therefore the higher predictive value of these scores.

Mr. **Weiss**. The interesting thing that I noted in the course of the hearing is that for the most part, people who directly deal with admissions and, therefore, with all aspects of the admissions criteria, and people who deal directly with tests, standard tests, tend generally to be supportive of opening the process up, except for those who are in the business of creating the tests, and for those who organizationally hire or contract with those who are in the business of making tests.

Why do you think there is that dichotomy between the established organizational entities and the individuals who deal within the field?

Dr. **DeWinter**. It is an organization that has existed for nearly 75 years and I think has undertaken many studies and developed a whole series of tests and patterns. I think there is a danger of becoming too convinced about our research and becoming too rigid in the way we view the results that we have. This is something which we are asking to take another look at and so perhaps it is a normal resistance, I would say, in anything that has existed for any number of years.

When something has reportedly been very successful and anyone challenges that, of course there is a certain amount of resistance to it.

My own feeling would be that it should stand the scrutiny and other people ought to be able to take a look. In some cases it may prove to do some things for certain groups of people and we ought to know that.

Mr. **Weiss**. Dr. Loewen, I was impressed by the testimony you gave that the problem of the bias that is built into the tests affects to a great extent not just minorities but rural students.

Could you expand on that?

Dr. **Loewen**. It is certainly very much the case in Mississippi, white or black, and not just because of its socioeconomic disadvantaged but also because of the rurality of the State.

In Vermont I experienced the fact that—and Vermont releases this data within its own confines—rural Vermonters, which means almost all, score much worse on the SAT exam than do those University of Vermont students who come from the suburbs of Boston, New York and Philadelphia, which make up half the student body. So we have a student body, half of whom make scores in the 500's and 600's, and the other half of whom make scores of 400, that come from Vermont. By the time they graduate they do the same quality work as those out-of-State students.

Mr. **Weiss**. Mr. Buchanan.

Mr. **Buchanan**. Thank you, Mr. Chairman. This is a very difficult subject for me; I am glad there is an agreement among us that there is an illness with the testing situation.

What I am not sure about is the instrument to deal with the illness. I am just not sure it is not good.

Let me be the devil's advocate and raise some questions some other people have raised for your response. For example, William Raspberry, in his usual namby-pamby way, said that this bill was strong, bitter stuff that doesn't secure anything, and he went on to quote, in his article in the Washington Post on September 10, Diane Ravage of Teachers College, Columbia University, who has this to say:

> It will create a bonanza for private tutoring services which would have actual tests for teaching students, not just their own inventions. This will increase the advantage of those who can afford to pay for coaching.

What would you respond to that?

Dr. **Franklin**. I don't agree with that, obviously, because in our testimony we indicated that there should be some regulating guidelines for coaching agencies in and of themselves. But there is also another issue involved which is that if coaching is effective—and I think that is a debatable issue that suddenly has been uncovered, that it is more effective than previously thought—then the kind of disclosure that this bill will afford will let the information, the principles and the general content of coaching, become available to a broader population than just coaching agencies.

Unless I am misunderstanding, all of a sudden we will have a new monopoly called testing coaching agencies. I feel that the bill will also have the effect of disclosing the general process of helping students to do better on tests and it will not become the exclusive right of exclusive coaching agencies.

Dr. **Newsom**. I am in total disagreement. I believe it will create coaching schools. I find it is done in cities and universities now. Sororities and fraternities do the whole business of coaching on tests. All of us know they do this. We used to have all the tests and get them to our friends.

Sure, they will be using these tests. If you know what the answers are going to be on the tests, you are going to teach it and work with it. If Raspberry said that I would agree with him completely. If you see the testimony you will see.

It happened when I was teaching at Southern University in Baton Rouge. We were behind in terms of the ROTC, we were making the lowest scores of all the schools basically in the South. We took the psychologist and the sociologist and we began to work on those things we knew were going to appear on those tests, and our scores were raised decidedly. So it is true.

Dr. LOEWEN. There seems to be the assumption underlying your question and underlying some of the testimony by Dr. Berendzen in particular, that to disclose a test would make it almost valueless and would raise the problem of just learning those questions.

It is the practice of ACT, I believe, and it is alleged to be the practice of ETS, to have a multiplicity of items on any test. Otherwise, my own teenage stepdaughter, who has taken the SAT last year, would score decidedly higher this year merely by remembering some of the questions she missed the first time and getting those answers right this time.

It is certainly claimed there are thousands of items on the SAT. It would seem obvious what you would disclose would be the test that is given next November and only those items that in fact went into the score as I understand the legislation.

Any testmaker—which I am and several of the rest of us—can make up a test which is a random selection or a specifically designated selection searched by the computer from the thousands of items you already have available and pull out some for the next exam. It will include some overlap with the items and it is perfectly easy to tell your computer in scoring to check the old versus the new.

If you have some student getting 100 percent of the old rate and 2.3 percent of the new, you flag that to somebody's attention or base the score primarily on the new items. It is no problem.

The gentleman from occupational testing where it may be a very specific test related to a specific field may have a reasonable disagreement to make to what I am saying. I am only saying this with regard to general aptitude testing of the type used for professional and graduate and baccalaureate schools.

Mr. BUCHANAN. Did you wish to comment?

Dr. DEWINTER. I would agree with the gentleman on the right. We seem to approach tests as if knowledge were an entirely finite sort of thing, that if you had a particular area there were only so many questions you could ask and if you knew that, that would be all there is to it.

We all know as teachers, too, we could give a test to a classroom and then after that say there were a few students who were absent, and somehow we can still make a valid test for those by asking the same principles under different circumstances.

We all remember those exams where you had to prove a theorem or an equation or an aspect in foreign language or anything like that, and you can give a very valid test and yet change the particulars of that particular test.

So I don't think it is that cut and dried.

Mr. WEISS. If you would yield for just one second, a couple of weeks ago, the New York Times in their magazine section of their Sunday newspaper, carried a feature story devoted to the coaching schools. I have a copy of this handy. They indicate that one of the schools, for example, has branches for courses given in 88 cities across the country. That is just one of many. There are a number of them across the country now.

It is not a question of coaching schools developing. I think the issue is they exist right now and perhaps if in fact we open the system up, those who learn from prior testing will not be just those who can afford it, too, for those coaches.

Thank you.

Dr. **Smith**. I would say in the occupational area, as I mentioned in my testimony, there are many licensing and certifying agencies who give their own tests in addition to the examination we might provide. I think we would be hard pressed under disclosure conditions to develop the necessary examination in the pool of items that would be required.

I think under those conditions they would not be able to meet their obligations. It would be of concern, obviously, to us as well, but it would probably be of lesser concern. But even so, as I mentioned in my testimony, a number of our clients have test candidates that number 30, 40, 50 candidates every year, and if we are talking about the development of a multiple-choice examination of 100 items or 160 or 200, which is about what we might use, the cross would be spread over many fewer candidates and, therefore, I think we are talking about some financial considerations here that are much greater than what would be anticipated in the aptitude area.

Also, if I may say, regardless of which bill one professes to support or reject, with respect to the credentialing systems in this country, they are extremely diverse and varied. I would caution, before the occupational clause in the Gibbons bill be considered in any other way for inclusion under whatever conditions, that there be much greater concern given to the nature of those systems before one passes that bill on to the social system.

Dr. **Franklin**. May I give an example of how the process can work as quickly as possible. In terms of coaching, we talk about coaching schools and how that is going to exclude more students, particularly minority students. We also have the problem in professional licensing of trying to give minorities licenses. I am involved with that.

There are professional schools to help professionals get their professional license. We have had black professionals that go to these professional schools and learn something about the system of coaching in order to take the examination. As a consequence, we have developed our own system of coaching and try to collect more information about the coaching process and the coaching game.

That can happen on a student level at an undergraduate school. You can dilute the impact that coaching agencies can have as more and more agencies begin to learn this is at their disposal. I think the NAACP would immediately get involved in helping its membership to learn how to help their students get coached on SAT and other examinations.

The other issue in terms of professional examinations, in New York State persons who do not pass the professional licensing examination in psychology—unless they have changed it since I took the examination—have the right to examine those questions they fail in the presence of an examiner. That was for both the objective part of the examination as well as for the essay part of the examination.

That may not be the best system but it is a way in which you can provide feedback for a professional or any person on their performance in a given testing situation. I think there are ways in which these systems can be devised and this bill helps to open up these kinds of options.

Mr. **Buchanan**. One more area of questioning. The Weiss bill would exempt admissions tests developed and administered by an individual school for its own purposes.

I wonder if such tests are likely to be more or less valid, reliable and subject to bias than the standardized test. Would anyone comment on that?

Dr. **Franklin**. I don't think we understood the question.

Mr. **Buchanan**. The Weiss bill exempts those tests given by an individual institution for its own purposes from the requirements that it lays down for standardized tests. I just wondered if there was a basis to assume those tests would be more valid, less biased, more reliable than the standard tests?

Dr. **Newsom**. I don't know if I understand completely what your question is. Let me see if I understand it at all.

One, I think you are asking would the standardized tests be more valid and reliable than would the test be that the institutions provide; or vice versa?

Mr. **Buchanan**. We are putting them in two categories, saying to the individual institution, you may continue to do your own testing and you are not required to disclose. I am asking if there is some basis for a double standard here so one step should be considered more biased, more reliable, more valid?

You see, we have two categories now: Individual school tests that do their thing and they don't have to disclose anything, and then we have the standardized test upon which we lay these requirements.

I am asking about the basis for this double standard.

Dr. **Newsom**. I would think that a standardized test that is used to exclude should be exposed to show the biases of that test. If the standardized test is used to exclude by a particular institution or occupation, I believe that should be exposed to the person—basically what was said a moment ago—but if it is a test that is used in the matter of showing progress or improvement or development, which is a different type test, achievement, then that kind of testing should not be.

In other words, if it would cause a lot of confusion.

I don't know if I really understand your question. I am probably not answering it very well.

Mr. **Buchanan**. Let me see if I can ask it a different way. Is there somehing to the idea that the individual school tests are likely to be less biased, more valid than the standardized tests?

Dr. **Franklin**. If that is the assumption, then it is an incorrect assumption.

Mr. **Buchanan**. They are excluded.

Dr. **Franklin**. I understand that.

Mr. **Buchanan**. That is what I am asking.

70

Dr. **Franklin**. There is always a problem of standardization and how valid and reliable a test is whether it is devised by individual institutions or by a nationwide testing agency. That is always a fundamental psychometric issue. That may be an area in which the bill might have to be reexamined. But I do know that in a larger system such as City University of New York, there is a combination—let me put it this way—using standardized tests with homemade individual tests. By that I mean for such areas as writing competency, there is more a tendency to try to evaluate written essays, and in the area of evaluating mathematical competency there is more tendency to rely on those standardized tests that measure mathematics.

But in either way there can be inherent biases in the assessment process.

Mr. **Buchanan**. Mr. Chairman, I will not press this further except to say I am really going to study all your testimony and the material from NAACP that is provided today, because I honestly don't know what we ought to do about this bill.

I know there is a problem, but one of the things that troubles me the most is the extent to which the tests may be accurate rather than inaccurate. If a child from age zero through six is allowed too much malnutrition, that child can be permanently handicapped for all of life. Perhaps not as dramatically but just as truly, if a child is subjected to too much discrimination in education and inadequate education in elementary and secondary school, that child can be educationally handicapped to the extent the child does not do as well on tests.

I just have to say, I grew up in a part of the country where I know that inadequate education was a problem for many years and still is a problem. There were some schools that were not good enough to teach high school students what they had to know in order to pass a test to enter a good college or pursue some other line.

I know there is a problem with testing. I am just saying that I am concerned about the extent to which these tests may be accurate rather than inaccurate.

Mr. **Steinbach**. I would like to address the questions. One, why the individual school tests. There are a number of reasons for this. One has to do with the number of students affected by any one examination. Second, the fact that predominantly examinations given by individual schools are only of a placement or achievement variety. A very small percent are of the SAT type.

A third reason has to do with the fact as I read the bill it seems not to deal with regulation of the conduct of colleges, their admissions offices or anything like that, and to move into the realm of the individual school would be to perhaps make that line somewhat shady.

I think they are all perfectly good reasons. You can only legislate so much at a time.

In response to the question posed that you are worried about the tests being right, I guess the answer is we can't really tell unless something like the disclosure legislation comes about. I couldn't tell. From what I can tell we know that even when you don't use any sophistication at all in terms of grade point average there is some objective on admissions applications. We know all these things have greater predictive validity as to how well a student will do in school, at least insofar as we have information at all than do SAT scores, for example.

But the SAT scores might do very well for some sets into some majors, say going into engineering. We don't know. We can't get access to the information.

Furthermore, if it does have high validity, it has to have it for the right reasons.

I will give you an example of wrong reasons. This was given to me by a law professor. What if we found out that students who liked strawberries make better lawyers? Is the question getting at that permissible on a SAT? My answer is "No." It has to have face validity.

How do students write that go to law schools? Do people who like strawberries come from an environment where they have learned to write and be able to afford strawberries? Obviously, I can make this into a case of accidental bias. It is slightly predictable for the wrong reason. We don't know.

Mr. **Buchanan**. Thank you very much.

Ms. Simmons said she just wanted to say amen to what somebody else said. It was such a resounding amen that the amen was stronger than the original statement.

I will have to take a very hard look at the testimony of the panel. I just hope you are right. I am afraid you are going to find there is a good deal else wrong.

Mr. **Goldstine**. I would like to comment, if I may. I came here today to illustrate how invalid tests can affect an entire industry.

The question has come up regarding the inability to determine whether or not an exam is valid. I do know that States around the country have been questioning the NCARB, which is the National Council of Architects Registering Board, since the implementation of its new exams back in 1973. NCARB has not responded and has only said its exam through its analysis is a valid exam.

In Wisconsin, a number of students questioned the current examination and they went to their legislature and they asked that this examination be validated. There was an abnormally high failure in that State as is around the country. Approximately 50 percent of the examinees—people who have 5 year degrees with 3 or more years of experience in the field.

They are now going into this with a background in the field and education before they sit for their examination. There is a prerequisite to have your internship and your education before you sit for the exam, and then there is a 50 percent failure rate.

The first study which was conducted by Dr. Glenn Taggetts from International Personnel Services, came out and said it doesn't seem as though this examination meets the guidelines of the APA or the EOC or the EEOC because a validation has never been conducted on this examination. They don't know whether or not it is valid.

Then there was a second study in which a group of architects was assembled in Wisconsin and 71 tasks were determined and there was a correlation done. There were a thousand questions on the exam and they found out the exam was approximately 50 percent related to the tasks that architects are doing in offices.

I think it is important that examinations, especially of this nature and of an academic nature, both be valid. People are putting numbers of years into decisions regarding professional entry to either the academic institutions or into their actual professions.

I think there are two issues here. There is the issue of minority discrimination which must be dealt with effectively but there is also the issue of overall professional discrimination against an entire group of applicants where an exam is designed to restrict entry into that profession not only because of minority reasons, which seems to be taking the brunt of these exams today, but also because of economic conditions in the country.

Architecture being a profession that closely relates to the economy at the time, it was one of the professions hard hit by the economic recessions of this decade. Their examinations seem to reflect that. I think that is an issue that really needs to be dealt with.

Mr. **Weiss**. Thank you very much. Mr. Erdahl.

Mr. **Erdahl**. Thank you.

Mr. Goldstine, I think you have already partially answered part of my questions and I appreciate your coming forth with a personal example. If some of my questions seem personal, I don't mean them that way.

How many people took that test and how many failed? Is this a problem nationwide? Is this similar to a bar examination?

Evidently you are indicating that the architectural profession wants to keep its group relatively small. Should the Government get involved in determining how many architects we should have or how many lawyers we should have?

You raise some very fundamental questions on who should be setting the criteria for getting into the profession. Should it be a group other than the profession itself? Is your case somewhat unique?

You say half the people who take the tests flunk it.

Mr. **Goldstine**. The failure rate is approximately 46 percent each year.

I brought this example up because I have done a great deal of research regarding this matter. I think the problem is that outsiders cannot regulate a profession. I believe that regulation should be left up to that profession itself. The issue is: If guidelines are not established by which these professions will test and allow for entry that will be on a fair basis, rather than to restrict entry—the trade offs here are to provide for public health safety and welfare and to allow for competition at a reasonable level that does not unreasonably constrain private enterprise.

The facts seem to indicate that a restraint of trade may be occurring.

I believe that a further investigation which is currently in process at this time will eventually come to a conclusion regarding these facts but, yes, it is a national concern, and States around the country have been requesting that NCARB respond.

Finally, after the exam was proved invalid in one State, and California is now being funded to replace the examination, I believe NCARB is beginning to respond.

What I question is unless some kind of specific guidelines are established, and evaluation can lead to proving an invalid exam valid based on the criteria for evaluation, I think it is time that guidelines be established, that professionals regulate themselves in a manner that is consistent with the overall objective of the Nation to allow the American dream to continue.

Thank you.

Dr. **Weiss.** Dr. Smith.

Dr. **Smith.** Since one of the basic issues has to do with the question of test validation—and it has also been raised with respect to the aptitude test as well—I think it should be stated that test validation is not necessarily an inherent part of a set of items. It is built into the procedures that are used to look at the items. So it is a process. It is a concept. And there is nothing inherent about one test that makes it necessarily valid versus another set of items.

It has to do with the conditions under which they are administered, taken, constructed, and so on.

This is why the standards for constructing psychological tests, the EO guidelines which currently exist covering the issue of test validity are not a cookbook procedure. It doesn't say take 10 items, crank them through formula 3 and come up with a test that is or is not valid. Validity is not a categorical decision. Data that are produced through the examination of a test need to be interpreted. Data do not speak. They need to be examined and clarified under the conditions under which the data were collected.

Therefore, I would say in response to the question, we need to look more carefully at the full range of validational procedures. I can't pass in judgment as to whether the test is or is not valid. This is a complex issue. But it is yet an issue that you are going to have to deal with in making the determination as to whether the bill has any worth.

What I am suggesting is that it is not a simple question. It is a complex psychometric question that needs to be examined here.

Mr. **Erdahl.** If I could ask Dr. Loewen to respond to an area of concern. You mentioned that educationally you are not opposed to tests but we need to have better tests. You mention the difficulty we sometimes find because of a localized situation, a rural situation, or the poverty area, that those people should perhaps have a different test.

Isn't there a danger in doing that to further isolate people or eliminate some challenging possibilities from them? For example, I grew up and went to a country school in Bush Creek Township in Minnesota. Maybe I would do well in a test oriented in an area like that but it wouldn't give me an opportunity where I might not do as well in a test that had a broader universe of concern.

Maybe I should have stayed in Bush Creek Township—but might you not also be expanding on the possibillity that people are going to be remaining in Tougaloo when they could go some other place or remain in a situation when they could do better if they went into some other type of activity or situation?

Dr. **Loewen.** I want to say I went to college in a small town in Minnesota so if they develop a Minnesota test, I want to take it along with you.

I think the point of my comment is not to suggest that we have alternate tests made up for people from rural areas or from the south or north. Rather, as I see the bill H.R. 4949, it requests that the testing agencies make available to the Commissioner, also ultimately to the Congress, data as to the regional variation. Actually, I suggest in that passage where it says race, age, sex and so on, that region be added. I think that region is a very important one and that rural status be added to the criteria of possible things that might correlate with test scores.

If those correlations were added so that we knew how blacks and whites scored, how people from rural Minnesota versus rural Texas scored, then one who is from Mississippi who takes the test and scores 365 on a 200 to 800 scale would have a possibility of scoring 800, would have the score interpreted in a little more broad context. In a sense a little more specific context, so it would be clear to anyone from Cornell Admissions that this student has done a dynamite job compared to other students from rural Mississippi. I don't know how Cornell knows that.

It is a very sophisticated school and maybe it does. Most schools cannot possibly put scores into context right now because the data is not available from ETS or from other testing agencies. Therefore, all they really know is the data of past people who have applied to other schools and they have a problem with the entire State of Mississippi. You see my point. I would just make this information available.

Mr. **Erdahl**. If I could follow up on that a bit, Mr. Weiss. You are suggesting that some other information be added to the test scores use for admissions, pertaining to rural/urban regions of the country. That is in addition to the testing process, quite independent of the bill before us today?

Dr. **Loewen**. I would suggest that the Weiss bill have added to it, and I make that suggestion specific; I think it's a very short phrase that would have to be added.

There is the phrase that the section requires "report to Congress the relationship between the test scores of test subjects and income, race, sex, ethnic, and handicap status."

I would suggest you simply add "regional and rural/urban" between the test scores and test subjects and income, race, sex, ethnic, rural/urban, and handicap status.

If that were not done, nonetheless, I still favor the bill, because it certainly requires much more openness and release of data than the test agencies purport to have in many cases.

Region and rural/urban are major concerns of mine.

Mr. **Erdahl**. One comment and then I will yield back my time. It seems some of that information would have been included on the application from the school. That information might be of interest to a sociologist doing some studies on the testing. All of that is pretty well included in the standardized application to an institution.

Dr. **Loewen**. Of course, the student who applies from Hollandale, Mo.——

Mr. **Erdahl**. Hollandale, Minn.

Dr. **Loewen**. OK; that student, Cornell or Yale or whatever, is aware of his residence and is probably at least vaguely aware that that is a rural area and it's in the South.

Mr. **Erdahl**. I would hope so.

Dr. **Loewen**. The point is that this section requires that average scores be released not only on the basis of race, on the basis of sex, and so on, but on the basis of region and rural ethnic composition.

That would be a very light burden because they already are corrected that way. They know where a test center is and if they would simply release the median score for Mississippi centers, then that would give Yale, as an admissions officer, or me as a sociologist, what a student who got a 385 and kind of wonders, am I stupid, that would give them a context to put their score within and a student from rural Mississippi would realize, knowing in their context they are doing pretty well and so would the school.

Mr. **Erdahl**. I understand what you mean.

No further questions, Mr. Chairman.

Mr. **Weiss**. Before we terminate the hearing for this morning, is there anyone who wants to submit additional written information?

If not, we express our appreciation.

Ms. **Simmons**. I would like to say NAACP would like, if you need additional information, we would be happy to try to provide some for the committee.

Mr. **Weiss**. We will probably take you up on your offer, but if anyone wants to submit additional written information for the record, we welcome that as well.

In closing I want to thank all of you for taking the time and trouble to come and spend a long and difficult morning with us.

Thank you very much.

[Whereupon, at 1:50 the Subcommittee on Elementary, Secondary, and Vocational Education of the Committee on Education and Labor adjourned.]

TRUTH IN TESTING ACT OF 1979; THE EDUCATIONAL TESTING ACT OF 1979

WEDNESDAY, OCTOBER 10, 1979

House of Representatives,
Subcommittee on Elementary, Secondary,
and Vocational Education,
Committee on Education and Labor,
Washington, D.C.

The subcommittee met, pursuant to notice, at 9:30 a.m., in room 2175, Rayburn House Office Building, the Hon. Carl D. Perkins (chairman of the subcommittee) presiding.

Members present: Representatives Perkins, Kildee, Weiss, Buchanan, and Erdahl.

Staff present: John F. Jennings, counsel; Jeffrey W. Brennan, staff intern; and Richard DiEugenio, minority legislative associate.

Chairman **Perkins**. The committee will come to order.

The Subcommittee on Elementary, Secondary, and Vocational Education is continuing hearings today on H.R. 3564 and H.R. 4949, two bills relating to standardized testing.

Today, in addition to hearing general testimony on these two bills, we will be focusing on two specific issues.

The first issue, which the General Accounting Office representatives will address, is whether the concepts of these two bills ought to be extended to cover tests administered by the Federal Government.

The second issue, which the National Education Association witnesses will address, is whether these "truth-in-testing" concepts should be extended to cover standardized tests given at the elementary and secondary levels and tests developed by teachers.

We have with us today Mr. Clifford Gould, Deputy Director, Federal Personnel and Compensation Division, U.S. General Accounting Office, and Mr. Donald Cameron, Assistant Executive Director for Programs, National Educational Association, accompanied by Mr. Frank Kovacs, Director of Research.

Now, come around, those gentlemen whose names I mentioned. We will hear the second panel at a later hour.

Come around, Mr. Gould and Mr. Cameron, and we will hear from you first, Mr. Gould.

We are delighted to welcome you here and you may commence. Without objection, your entire testimony will be inserted in the record, and if you want to summarize it, fine, go ahead.

STATEMENT OF CLIFFORD GOULD, DEPUTY DIRECTOR, FEDERAL PERSONNEL AND COMPENSATION DIVISION, U.S. GENERAL ACCOUNTING OFFICE, ACCOMPANIED BY JOHN HARPER, PROJECT MANAGER; DONALD CAMERON, ASSISTANT EXECUTIVE DIRECTOR FOR PROGRAMS, NATIONAL EDUCATION ASSOCIATION, ACCOMPANIED BY FRANK KOVACS, DIRECTOR OF RESEARCH

STATEMENT OF CLIFFORD GOULD, DEPUTY DIRECTOR, FEDERAL PERSONNEL AND COMPENSATION DIVISION, U.S. GENERAL ACCOUNTING OFFICE

Mr. GOULD. Good morning.

I would like to introduce Mr. John Harper, who is a personnel psychologist who heads our Management Science Group within the Division.

We appreciate the opportunity to appear before you today to discuss the potential impact of extending concepts contained in H.R. 3564 and H.R. 4949, to tests administered by the Federal Government.

You asked that we focus on our report, "Federal Employment Examinations: Do They Achieve Equal Opportunity and Merit Principle Goals."

Our report, released on May 15 of this year, highlighted problems confronting those responsible for insuring that the Federal work force is competent, productive, and to the extent possible, reflective of all segments of our Nation's society. The goal of Federal employment programs is to hire qualified individuals based on merit principles giving everyone an equal opportunity to compete for a job.

To achieve this goal the Civil Service Commission, the predecessor agency to the Office of Personnel Management, designed various examining procedures to help assure that the most competent and productive people are employed.

We found that black job applicants were screened out of the competition at a much higher rate than white applicants on two major written tests used by OPM to develop hiring registers. The implications of this are presented in detail in our report where we recommended that OPM further investigate the use of valid alternatives to written tests.

Mr. Chairman, I would like to include for the record those recommendations.

Chairman PERKINS. Without objection, it is agreed to.

Go right ahead.

Mr. GOULD. I would like to begin my comments on H.R. 3564 and H.R. 4949 with a caveat. In the time available to prepare this testimony, we were unable to perform a detailed legal analysis of these two bills. We were also unable to audit cost and workload estimates provided to us by OPM. We can provide additional information for the record at a later date or the subcommittee may want to obtain that information directly.

While we support the intent of both H.R. 3564 and H.R. 4949 with respect to improving the quality and availability of information about tests, we do have reservations about the potential cost and feasibility of extending certain provisions to Federal examining.

The Federal Government may well be at the forefront in providing information and preparatory assistance to those taking standardized written tests. For instance, in 1971 it published a 60-page booklet designed to provide teachers, students, and other interested persons with samples of the types of items used in Government tests for filling vacancies in a wide variety of office occupations.

This booklet has undergone five revisions over the years and each time it has been revised OPM produced 10,000 copies for distribution to those who might be interested. In addition, it has been available through the Superintendent of Documents.

In 1974, the Government published a 12-page booklet containing samples of the types of items used in the PACE written test. This booklet was attached to the basic PACE qualifications announcement in 1975 and has been distributed to every applicant since then.

The most important reservations we have about these bills relate to section 5 of H.R. 4949, requiring the public disclosure of all test items, and to that part of section 6 in H.R. 3564 which prohibits the assignments of scores to knowledge and achievement tests on the basis of rank order.

My statement contains a rather detailed discussion as to why we have objections to those, and I will not go into them here. In addition to the comments we have provided on rank ordering, we would like to point out the veterans' preference legislation does require a rank ordering of applicants' and some modification of that legislation would be necessary if this bill were to be enacted as it now stands.

I would like to conclude by repeating that we support the intent of these bills to improve and make public the processes and information so important in enjoying an equal chance for achievement in our society.

The hearings of this subcommittee and the legislation it is considering have added materially to public scrutiny and accountability in this area. With respect to those provisions about which we have some concern, we believe that their enactment might well curtail testing in the Federal Government. As I mentioned earlier, we have recommended that OPM further investigate the use of valid alternatives to written tests.

This completes my prepared statement. We will be happy to answer any questions you may have.

[Prepared statement of Clifford I. Gould follows:]

PREPARED STATEMENT OF CLIFFORD I. GOULD, DEPUTY DIRECTOR, FEDERAL PERSONNEL AND COMPENSATION DIVISION, U.S. GENERAL ACCOUNTING OFFICE

We appreciate the opportunity to appear before you today to discuss our recent report on Federal employment examining and to comment on the potential impact of extending concepts contained in H.R. 3564, the Truth in Testing Act of 1979, and H.R. 4949, the Educational Testing Act of 1979, to tests administered by the Federal government. You asked that we focus on our report, "Federal Emloyment Examinations: Do They Achieve Equal Opportunity and Merit Principle Goals."

Our report, released on May 15th of this year, highlighted problems confronting those responsible for ensuring that the Federal work force is competent, productive, and to the extent possible, reflective of all segments of our Nation's society. The goal of Federal employment programs is to hire qualified individuals based on merit principles giving everyone an equal opportunity to compete for a job. To achieve this goal the Civil Service Commission, the predecessor agency to the Office of Personnel Management, designed various examining procedures to help assure that the most competent and productive people are employed.

We found that black job applicants were screened out of the competition at a much higher rate than white applicants on two major written tests used by OPM to develop hiring registers. The implications of this are presented in detail in our report where we recommended that OPM further investigate the use of valid alternatives to written tests.

The written test for the Professional and Administrative Career Examination (PACE) is the most thoroughly researched test in the history of Federal civil service examining. According to OPM the test has been fully validated and its research shows a clear relationship between performance on the test and subsequent performance on the job. OPM says that not only does the test result in hiring individuals who perform better on the job, but its use enhances overall Government productivity.

The PACE is particularily important because it is the entry route into Federal service for more college graduates than any other single method and because the type of jobs covered by the test often leads to a higher level career position. For that reason, it is important that the Subcommittee explore the potential impact of the bills it is considering on this examination.

I would like, first, to put PACE in perspective in terms of its place in obtaining a Federal job. PACE is used to hire new employees for entry level jobs in 118 different occupations. According to OPM, the most common feature of PACE occupations is their information burden--that is, all of the occupations require individuals who can read, understand and interpret a large body of written laws and regulations or a large body of knowledge related to a specific profession or specialty area. OPM determined that five abilities were important to successful performance in these occupations and could be measured with a written test. These were: inductive and deductive reasoning, verbal comprehension, judgment, and quantitative reasoning.

Individuals who pass the test and receive a rating of 70 or more are listed on a PACE job register according to their final rating--test score plus points for Veterans' preference or for being an outstanding scholar, when applicable. When entry level jobs become available in one of the 118 occupations, agencies request OPM to send them a list of names from the PACE register. In response, OPM sends the agency the three names with the highest ratings. Agencies generally must select one of the three individuals on the list, and if one of the individuals on the list is an eligible Veteran, the Veteran must be selected. The process is somewhat more complex than what I have just described and substantial changes are contemplated as a part of the authority delegations under the Civil Service Reform Act. But this should help explain what has been required of an individual seeking an entry level PACE position and who has not had a Government job.

During fiscal year 1978, OPM processed 1.6 million job applications for all types of competitive civil service jobs, and about 152,000 people were hired. These numbers include all competitive jobs, both wage grade and general schedule. During that same year 135,000 individuals took the PACE, 77,000 passed with a score of 70 or above, and 7,600 were hired from PACE registers.

COMMENTS ON EXTENDING CONCEPTS IN H.R. 3564 AND H.R. 4949 TO FEDERAL EXAMINING

I would like to begin my comments on H.R. 3564 and H.R. 4949 with a caveat. In the time available to prepare this testimony, we were unable to perform a detailed legal analysis of these two bills. We were also unable to audit cost and workload estimates provided to us by OPM. We can provide additional information for the record at a later date or the Subcommittee may want to obtain that information directly.

While we support the intent of both H.R. 3564 and H.R. 4949 with respect to improving the quality and availability of information about tests, we do have reservations about the potential cost and feasibility of extending certain provisions to Federal examining.

The Federal government may well be at the forefront in providing information and preparatory assistance to those taking standardized written tests. For instance, in 1971 it published a 60-page booklet designed to provide teachers, students, and other interested persons with samples of the types of items used in government tests for filling vacancies in a wide variety of office occupations. This booklet has undergone five revisions over the years and each time it has been revised OPM produced 10,000 copies for distribution to those who might be interested. In addition, it has been available through the Superintendent of Documents.

Again, in 1974 the government published a 12-page booklet containing samples of the types of items used in the PACE written test. This booklet was attached to the basic PACE qualifications announcement in 1975 and has been distributed to every applicant since then.

The most important reservations we have about these bills relate to Section 5 of H.R. 4949, requiring the public disclosure of all test items, and to that part of Section 6 in H.R. 3564 which prohibits the assignment of scores to knowledge and achievement tests on the basis of rank order.

We argree with the various arguments which have been made to the effect that the Federal government would experience a substantial economic burden if it were required to provide copies of test questions and correct answers to examinees. Such a procedure would preclude reuse of any items. So new ones would have to be prepared each time a test is given. This has enormous cost and technical implications.

The cost implications are illustrated in an example prepared by OPM. The government's clerical exam is administered nationwide every working day. If test items were disclosed, 250 separate and unique exam booklets would be needed each year. It takes about one staff year, at an average salary of $20,000, to prepare a clerical exam booklet using existing questions. OPM estimated that at least three staff years would be required to prepare a booklet using totally new questions. This does not include additional administrative costs required to insure that the same test is given nationwide on each day so that applicants in one area do not benefit from disclosure in another area.

The administrative problems, budget, and staff requirements are multiplied many times when more complicated, higher level exams are used. To develop the current pool of items for the PACE, for instance, required a total of 52 staff years of effort over a 5-year period.

It should be noted that the cost and workload estimates I have just provided are for test development only. They do not cover the very time-consuming and costly process of validating the test once it is developed.

The technical implications are no less severe. In the first place, there is a finite limit to the number of new questions that can be developed in any subject matter area without quality deteriorating to a level that the exam would no longer be an effective means of evaluating people for employment. In the second place, a group of common questions is needed in each version of a test so that adjustments can be made for differences among the versions in average difficulty. Only in this way can a score have the same meaning on one version as it has on another. Disclosure would make it impossible to have a common group of questions. Therefore different versions of the same test could not be equated and changes over time in test performance--such as the decline in SAT scores--could not be identified.

The concern we have with H.R. 3564 is with the distinction made in Section 6(c) between aptitude test and achievement test and the prohibition against grading the latter on the basis of the relative distribution of scores of other test subjects.

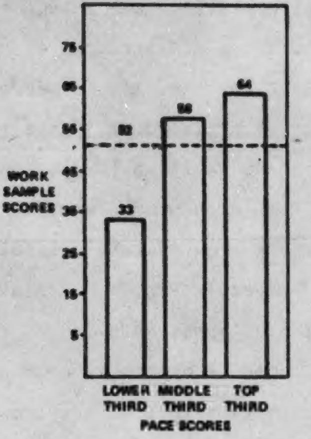
The distinction between aptitude and achievement tests is a point of great controversy even among experts in the field of testing and assessment. Sometimes the distinction is made on the basis of the way the test scores are used: if they are used to assess someone's current level of performance, the test might be considered an achievement test; if the scores are used to predict how well someone might perform in the future, the test might be called an aptitude test. By this distinction, the clerical exam I mentioned earlier would seem to be an achievement test, while the PACE would seem to be an aptitude test.

Whether or not this type of distinction is made, we would be concerned about the extension of a prohibition such as described in Section 6(c) to Federal examining. We would be particuarly concerned about extending the prohibition to aptitude tests like PACE; but we would also be concerned about extending it to achievement tests like the clerical exam where rank order on the exam may be related to relative performance on the job. The reason for our concern can be seen in data on the value of PACE reported in our study.

As soon as the PACE written test was developed, OPM started research designed to determine its value in terms of the extent to which scores on it were related to job-performance in a number of occupations. The first study was conducted on social security claims examiners. The written portion of PACE was given to 250 examiners and measures of their work effectiveness were collected. These measures were specially designed for the study. They included a work sample consisting of a standardized claim to be adjudicated by the examiner and scored for correctness.

OPM's analysis of the relationships between test scores and the various job performance measures showed that those who scored higher on the test tended to be the ones who also scored higher on the measures of job performance. The graph on the following page gives some indication of these relationships. It shows that those who scored in the top third on PACE did almost twice as well on the work sample (average score of 64) as those who scored in the bottom third on PACE (average score of 33).

Average Work Sample Scores at Each Range of PACE Scores



What the chart shows is that the average Work Sample Score for all three ranges of PACE scores combined is 52. That is the score you would expect a large group of people to obtain on average if they were selected on some basis unrelated to their relative performance on PACE. However, there is a relationship between performance on PACE and performance on the Work Sample. So if you want to maximize productivity you would select only those who score in the highest third on PACE. Their Work Sample scores average out to 64--an improvement of 12 points over selection among a group which has not been rank-ordered.

Another way to look at this is in terms of how much productivity is increased by selecting on the basis of relative rank order as opposed to selecting on some other basis. OPM researchers have constructed economic models which show that for PACE the increase in productivity is over twice as high for selection done in a strict rank-order based on test scores as it is for selection done on a random-order basis from among those who pass the test.

In addition to these productivity considerations, it should be noted that Veteran's Preference legislation does require a rank ordering of applicants.

Mr. Chairman and Members of the Subcommittee, I would like to conclude by repeating that we support the intent of these bills to improve and make public the processes and information so important to enjoying an equal chance for achievement in our society. The hearings of this subcommittee and the legislation it is considering have added materially to public scrutiny and accountability in this area. However, with respect to those provisions about which we have some concern, we believe that their enactment could bring about the end of testing in the Federal government.

This completes my prepared statement. We will be happy to answer any questions you may have.

COMPTROLLER GENERAL'S
REPORT TO THE CONGRESS

FPCD-79-46

FEDERAL EMPLOYMENT EXAMINATIONS: DO THEY ACHIEVE EQUAL OPPORTUNITY AND MERIT PRINCIPLE GOALS?

RECOMMENDATIONS TO THE DIRECTOR, OFFICE OF PERSONNEL MANAGEMENT

A comprehensive review of the Federal policies and practices affected by the new Uniform Guidelines is being planned by the Office of Personnel Management. This review will include job qualification standards, examinations, performance appraisals, and promotion policies.

GAO endorses this review. However, actions are needed which cannot await the outcome of the planned review. The Director, Office of Personnel Management, should:

--Increase efforts to comply with the Uniform Guidelines requirement for collecting and maintaining records which will disclose the impact of tests and other selection procedures on the employment opportunities of minority group members and females. As a part of the recordkeeping system required by the Uniform Guidelines, the Office should design a system to track, over a period of time, the job performance of groups of people selected from job registers and those selected using other methods. If the tracking system shows a selection procedure that has less impact on minorities and results in selecting employees who perform as well or better than those selected from the registers, then the use of that procedure should be expanded. If the tracking system shows that employees selected by an alternative method perform less well than those selected from registers, then the alternative should be reexamined and a decision made concerning its continued use.

--Reevaluate the PACE validation strategy to assure that it conforms with the Uniform Guidelines. In the interim, assure that the planned comprehensive job analyses of the remaining occupations are completed.

If further validity studies are necessary, GAO recommends that the Director:

--Reconsider using the predictive design or consider performing a research study in which all applicants selected for a PACE occupation by some other merit procedure are administered the test and followed up at a later date.

--Assure, to the extent possible, that research participants are representative of the relevant labor market.

--Investigate the fairness of the test for race, sex, and ethnic subsamples.

Chairman **Perkins**. Thank you.

We will now hear from you, Mr. Cameron. Go right ahead.

STATEMENT OF DONALD CAMERON, ASSISTANT EXECUTIVE DIRECTOR FOR PROGRAMS, NATIONAL EDUCATION ASSOCIATION

Mr. **Cameron**. I am Don Cameron, assistant executive director of the National Education Association.

With me is Mr. Frank Kovacs, who is the director of research for the NEA and available to help if anybody has any questions.

The National Education Association is pleased to have been invited to discuss the Truth in Testing Act of 1979, H.R. 3564, and the Educational Testing Act of 1979, H.R. 4949, which would be extended to include standardized tests given at the elementary and secondary level, and also teacher-prepared tests.

On August 1, 1979, when NEA presented testimony on H.R. 4949 before the Committee on Education and Labor, it supported the bill and urged the Congress to proceed and to expedite the passage of H.R. 4949.

At this time NEA reconfirms its endorsement of H.R. 4949.

The Association believes passage of this legislation will insure that standardized tests will not continue to deny equal educational opportunity for all people.

Both H.R. 3564 and H.R. 4949 represent notice-and-disclosure legislation; however, they differ substantially on the types of tests covered, the extent of involvement of the Commissioner of Education, and the type of enforcement provisions.

Furthermore, for the most part, the disclosure requirements of H.R. 3564 require the same type of information currently being provided voluntarily by testing agencies.

Because H.R. 3564 does not require disclosure of underlying data on the examinations, it would not enable professionals outside the testing industry, including teachers, to analyze or comment on test construction and validity.

Moreover, H.R. 3564 fails to provide for the disclosure of scoring data in addition to the disclosure of test scores that may be given to educational institutions.

NEA encourages the committee to use H.R. 4949 as the basis for the proposed testing legislation and incorporate H.R. 3564 as part of the proposed legislation where it is deemed appropriate.

In relation to the questions concerning the impact of the proposed legislation at the elementary and secondary level and to teacher preparedness tests, the NEA does not recommend the extension of H.R. 4949 to the elementary and secondary level at this time.

The proposed legislation currently addresses tests used at the postsecondary level for admission selection. Testing experts generally classify this type of test as an aptitude tests. Examples of aptitude tests are the American college testing program, scholastic aptitude test, law school admissions test, medical college admissions test, dental admissions tests, and the Miller analogies test.

The common element here is that these tests are used to predict future success in college course work. Aptitude tests usually are not designed to predict a student's future success in a chosen field, nor do the developers of these tests make such a claim.

The NEA believes that the estimated 2.5 million test takers at the postsecondary level already constitute a large enough group to determine how well the Federal legislation will work with aptitude tests.

By extending the legislation to include elementary and secondary students, the number of test takers affected would exceed 47 million, and the additional types of tests included would be achievement, ability, and IQ tests.

The implications of the increase in both the number of test takers and the types of tests cannot be adequately assessed at this time. In addition to the implications for testing companies, school district costs for testing the logistics of administration, legislation enforcement cannot be accurately determined.

The association further believes that the state of the art within the testing industry is in need of a major overhaul. In NEA's August 1, 1979, testimony before this committee, a segment of a speech made by Oscar Buros was quoted regarding problems and needed changes in testing.

In that same speech, Buros outlined a program for improving testing practices. Buros' program has been available to test publishers for over 30 years. Unfortunately, they have not taken his sound expert adivce about how to improve test measurement.

The association believes that the Buros proposal for improving tests and their use should be implemented. His proposal would help realize NEA's position on achievement testing. We strongly support the use of student achievement testing to diagnose learning needs, to prescribe instructional activities, and measure student progress in the curriculum content taught by the classroom teacher.

The association believes that such achievement testing is the responsibility of the local school district and teachers at the building level.

Moreover, NEA opposes the use of standardized achievement testing for the purpose of evaluating school programs because of the belief that the National Assessment of Educational Progress, NAEP, provides a more appropriate approach to assessing school programs.

NEA believes that the state of the art in the field of testing needs to be totally assessed at all levels. The proposed legislation, H.R. 4949, as currently structured, begins this process at the postsecondary level. The extension of the legislation to the elementary and secondary school level should be postponed at this time.

Finally, Mr. Chairman, NEA respects and values the time of the members of this committee. Therefore, NEA research has prepared a technical supplement to this testimony for use by the committee.

This supplement defines and develops technical concepts and issues which NEA believes must be recognized before further action to extend this vital legislation beyond the postsecondary level is implemented.

NEA Research would like to extend any needed technical assistance or provide additional information that this committee may deem appropriate.

Thank you very much.

[The prepared statement of Donald Cameron follows:]

PREPARED STATEMENT OF DON CAMERON, ASSISTANT EXECUTIVE DIRECTOR, NATIONAL EDUCATION ASSOCIATION

FOREWORD

The National Education Association has presented testimony on the Truth-in-Testing Act of 1979 (H.R. 3564) and the Educational Testing Act of 1979 (H.R. 4949) on two occasions, August 1, 1979, and today, October 10, 1979. The Association supports the concept of this type of legislation primarily because the testing industry is in need of a major overhaul.

The NEA proposes a course of action which would make testing and human measurement in this country consistent with the ideas proposed by Oscar Buros and Ralph Tyler. The National Assessment of Educational Progress (NAEP) and other criterion-reference measures are models for the testing industry. The NAEP approach, if properly designed, implemented, and used, would move the testing industry from perpetrators of a differentiation-type of measurement to a model of measurement that would provide information about course and learning objectives.

This technical supplement provides selected information about testing and measurement related to the proposed legislation.

NEA Research is in the process of preparing a research memorandum on testing and measurement. The memorandum is designed to address systematically many of the concepts discussed in this technical supplement. The three sections to this supplement are (1) the text, (2) Appendix A: NEA's Analysis of H.R. 3564 and H.R. 4949, and (3) Appendix B: NEA's letter of support for the National Assessment of Educational Progress.

NEA's TECHNICAL SUPPLEMENT FOR THE STATEMENT ON THE EDUCATIONAL TESTING ACT OF 1979

October 10, 1979

The National Education Association is pleased to have been invited to discuss the Truth in Testing Act of 1979 (H.R. 3564) and the Educational Testing Act of 1979 (H.R. 4949), which would be extended to include standardized tests given at the elementary and secondary level and also teacher-prepared tests.

When NEA presented testimony on H.R. 4949 before the Committee on Education and Labor on August 1, 1979, it supported the bill and urged the Congress to proceed and expedite the passage of H.R. 4949. At this time NEA reconfirms its endorsement of H.R. 4949. The Association believes passage of the legislation will ensure that standardized tests and/or the inappropriate use of standardized tests do not continue to deny equal educational opportunity for all people.

Both H.R. 3564 and H.R. 4949 represent notice-and-disclosure legislation; however, they differ substantially on the type of tests covered, the extent of involvement of the Commissioner of Education, and the type of enforcement provisions. Furthermore, for the most part, the disclosure requirements of H.R. 3564 require the same type of information currently being provided voluntarily by testing agencies. Because H.R. 3564 does not require disclosure of underlying data on the examinations, it would not enable professionals outside the testing industry, including teachers, to analyze or comment on test construction and validity. Moreover, H.R. 3564 fails to provide for the disclosure of scoring data in addition to the disclosure of test scores that may be given to educational institutions.

NEA would encourage the Committee to use H.R. 4949 as the basis for the proposed testing legislation and incorporate H.R. 3564 as part of the proposed legislation where it is deemed appropriate.[1]

---

[1] See Appendix A for NEA's analysis of H.R. 3564 and H.R. 4949.

Impact of H.R. 4949 at the Elementary and Secondary Level and on Teacher-Prepared Tests

In relation to the questions concerning the impact of the proposed legislation at the elementary and secondary level and to teacher-prepared tests, the *NEA does not recommend the extension of H.R. 4949 to the elementary and secondary level at this time.*

The proposed legislation currently addresses tests used at the post-secondary level for admission selection. Testing experts (psychometricians) generally classify this type of test as an aptitude test. Examples of aptitude tests are the American College Testing Program (ACT), Scholastic Aptitude Test (SAT), Law School Admission Test (LSAT), Medical College Admission Test (MCAT), Dental Admission Test (DAT), and the Miller Analogies Test (MAT).

The common element in all is that these tests are used to predict future success in college coursework. Aptitude tests usually are not designed to predict a student's future success in a chosen field (nor do the developers of these tests make such a claim).

The NEA believes that the estimated 2.5 million test takers at the post-secondary level constitute a large enough group to determine how well the federal legislation will work with aptitude tests. By extending the legislation to include elementary and secondary students the number of test takers affected would exceed 47 million and the additional types of tests included would be achievement, ability, and IQ tests. The implications of the increase in the number of test takers and types of tests cannot be adequately assessed at this time. In addition to the implications for testing companies, school district costs for testing and the logistics of administration and legislation enforcement cannot be accurately determined.

The Association further believes that the state of the Art within the testing industry is in need of a major overhaul. In NEA's August 1, 1979 testimony before this committee a segment of a speech made by Oscar Buros was quoted regarding problems and needed changes in testing. In that same speech Buros outlined a program for improving testing practices. Buros' program has been available to test publishers for over 30 years. Unfortunately they have not taken his sound expert advice about how to improve test measurement.

The Association believes that the Buros proposal for improving tests and their use should be implemented. His proposal would help realize NEA's position on achievement testing. He proposed the following:

> I suggest that achievement test batteries be of two types: tests for assessing the performance of groups and tests for assessing the performance of individuals. I shall refer to these two categories as group tests and individual tests where the words "group" and "individual" refer to the targets being evaluated.
>
> The group tests should be designed to measure the achievement of schools, school systems, or other groups having common objectives and learning environments. Since our interest is in groups rather than individuals, each test could be quite short, requiring very little time to administer. In situations where the groups are fairly large, the tests could be subdivided into much shorter subtests with no student taking more than one subtest. The time now required to administer an achievement battery, sometimes as much as seven hours, could be reduced to thirty minutes. With group tests such as these, schools would be able to compare their performance on item, part, and total scores as well as their frequency distributions with those of normative groups. In addition to covering objectives common to most school systems, these group tests could be supplemented by subtests covering objectives not now included in standardized tests. The use of short group tests, each taken by only a fraction (say, one-fifth) of the students will greatly reduce the costs in time and money of testing. It will stimulate test publishers to give greater attention to problems involving group measurement and the interpretation of group results. A much wider range of objectives and curricular analyses could be covered. Relatively simple norms, raw scores means, frequency distributions of raw score means, both for total scores and item scores, would replace the complicated systems of norms now provided for individuals. I would not, however, limit the normative data to the particular grade for which the test was designed. The same normative data should be presented for the grades immediately below and above the target grade. The presentation of this comparative information for adjacent grades is especially important in the interpretation of tests built for measurement rather than for differentiation.
>
> As is true of all achievement tests, scores on the group tests would be not only a function of what a student has learned, but also a function of the closeness with which the test covers the instructional objectives and curricular offerings of a local school system. Although these learning and curricular effects are confounded, I would like to experiment with ways of considering each of the effects separately. Let me suggest one way in which this might be done.

> Let all normative groups provide information on the extent to which the test specifications and the test items are appropriate locally. Checklists could be provided requiring school officials or committees of teachers to classify each objective and each item into four categories for each grade: (1) taught in an earlier grade, (2) taught in the target grade, (3) taught in later grades, and (4) not taught in any grade. Item score norms might be presented for each of these categories. Other questions might well be directed to the importance of an objective or item. The normative groups might also be asked to indicate local objectives not covered by the group tests. Statistics such as these would enable school systems to compare both the performance of their students and the extent to which the objectives and content of the tests coincided with their local objective and instructional programs.
>
> The use of different tests for measuring groups and individuals would permit school systems to abandon national norms altogether for individuals and to adapt and supplement commercially purchased tests and processing services to better meet local needs. Local norms would be far more meaningful than national norms. I would use compound scores consisting of a percentile rank within grade and the raw score (corrected for guessing) as a percentage of the possible score. Purchased tests could be supplemented by locally prepared examinations and integrated into the testing program.
>
> Since there would be no national norms and statistical procedures for selecting items would not be used, the costs of final tryouts and normative testing would be eliminated. Local school systems would be free to adapt the tests in various ways to better meet their needs. Items could be dropped by not scoring and new items added in locally constructed tests. Time limits and administration procedures could be changed if desired. School systems would become more actively involved in the choice, study, adaptation, and supplementation of commercially purchased tests and processing services. Test authors and publishers should give local school systems assistance in formulating testing programs which are adapted to the local situation.
>
> By better meeting local needs, publishers are likely to sell more tests to serve as a flexible core of the local testing program. Since computers would be necessary to provide local norms and reports for both students and test items, the data processing services of publishers would be needed more than ever.
>
> I shall not take the time to enumerate further the advantages of using national norms for group measurement and local norms for individual measurement. If a school system were forced to choose only one of these two kinds of tests, I would have no hesitation in recommending the locally adapted and supplemented test with local norms.[2]

Buros' program of improvement for test construction and use has been available to test publishers for over 30 years. There has been little improvement in the direction he proposed.

---

[2] Buros, Oscar K. "Fifty Years in Testing: Some Reminiscences, Criticisms, and Suggestions." *Educational Researcher* 6: 9-15, July-August 1977.

Test publishers and the industry at large continue to cling to the 1920s concept of testing.

Prior to NEA's examination of Buros' concepts and testing at the elementary and secondary level the Association would like to discuss briefly the following:

- Type of test addressed by the proposed legislation (H.R. 4949)
- Types of tests
- Use of tests
- NEA's position on the types of tests
- Use of tests in reaching decisions
- Criteria for evaluation of tests

Type of Test Addressed by the Proposed Legislation (H.R. 4949)

Section 10 of H.R. 4949 defines the term *standardized test* or *test* as meaning:

> Any test that is used, or is required, for the process *of selection for admission* to postsecondary educational institutions or their programs,
>
> or
>
> Any test used for preliminary preparation for any test that is used, or required, for the process of *selection for admission* to postsecondary educational institutions or their programs, which affects or is conducted or distributed through any medium of interstate commerce, but such term does not include any test designed solely for nonadmission placement *or credit-by-examination* or any test developed and administered by an individual school or institution for its own purposes only. (Emphasis added.)

The critical concept embraced by the proposed legislation as previously stated is on "selection for admission," and is considered to be an aptitude test. It is used to forecast or predict success.

Unfortunately, the lack of forecasting ability (predictive validity) of the tests with regard to on-the job success does not prevent admissions offices from using this type of test as part of the pre-service selection process for various occupations. This not only raises the question of testing for forecasting or predicting results, it also moves the focus from use of inadequate aptitude tests to the inappropriate use of test results; namely, a student's admission to a college or a graduate program.

In summary, the type of tests that the Educational Testing Act of 1979 (H.R. 4949) addresses are Aptitude Tests. This type of test has been used for classifying students, assigning students for instruction (ability grouping) and in some instances as a substitute for IQ tests. The principal use of the aptitude test is to predict future performance (e.g., first semester college grades).

Types of Tests

Most psychometricians classify tests into two categories, Maximum and Typical performance tests.

- *Maximum Performance Tests* are designed to measure a person's best performance. Intelligence, Aptitude and Achievement tests are classified as Maximum Performance Tests.
    –*Intelligence Tests* generally are designed to measure verbal, nonverbal, memory and problem solving. Some psychometricians define intelligence as what the test measures. Intelligence tests are for most purposes thought of as general or scholastic aptitude.
    –*Aptitude Tests* are theoretically designed to measure mental operations that improve little with practice and provide the basis for predicting future level of performance.
    –*Achievement Tests* are designed to measure skills, knowledge and competence. This type of test measures the degree of accomplishments in a training or educational experience.
- *Typical Performance Tests* are designed to measure how a person reacts, feels or behaves. Personality tests, interest inventories, situational tests and projective techniques are illustrations of Typical Performance Tests.

The purpose in measuring typical performance is to determine how a person behaves in daily activities. These types of tests are designed to measure how a person would normally behave, react or feel in a given situation. The assumption underlying this type of measurement is that the test taker is responding honestly to questions about behavior and/or feelings.

731

Use of Tests

Tests are used in education for research, prognosis, and diagnosis. Prognosis or prediction and diagnosis are the two most frequent uses of tests. In *prediction*, the emphasis is on differences between an individual's performance and that of other individuals or a standard. *Diagnosis*, on the other hand, focuses on the analysis and description of an individual's performance from task to task.

Aptitude tests and I.Q. tests are examples of tests that are used for prediction. They are used to differentiate among people taking the tests and to predict future performance. Achievement tests when used to measure a person's skills or knowledge (not competence) are diagnostic and provide a means to isolate those characteristics which need to be corrected (e.g., difficulty in comprehending complex sentences).

NEA's Position on Achievement and Aptitude Tests

NEA strongly supports student achievement testing to diagnose learning needs, prescribe instructional activities, and measure student progress in the curriculum content taught by the classroom teacher. The Association believes that such achievement testing is a responsibility of the local school district and teachers at the school building level.

Further, the Association believes that the National Assessment of Educational Progress (NAEP) provides a more appropriate approach to assessing school programs and student competencies. The NAEP program, in operation since 1964, has emerged as an accepted method to assess educational programs by both teachers and administrators. NEA supports the NAEP program as the most appropriate approach for assessing educational programs. (See Appendix B, letter of support to NAEP.)

Aptitude tests imply prediction and are used to forecast future performance levels. They have been used for selecting or excluding individuals for admission to college, graduate school, and jobs.

NEA opposes the use of tests that deny students full access to equal educational opportunities.

Aptitude tests and those achievement tests which are designed to measure competence and are used to deny students full access to an education or future employment are the type of tests that NEA would like to see eliminated from use.

Use of Tests in Reaching Decisions

Tests are used for two types of decisions, *institutional* and *individual*. Institutional decisions are decisions made on behalf of an institution such as a college or school in selecting or rejecting an applicant. A bad decision about an individual tends to have little if any adverse effect on the institution. Furthermore, tests predict group performance better than individual performance.

Individual decisions are decisions about a particular person (e.g., daughter, son). The emphasis in this type of decision making is on the individual and not on the category of people with the same or similar characteristics. An incorrect decision may have a life-long effect.

Institutional decisions generally are made in a centralized setting while individual decisions tend to be made in a decentralized setting. For example, an admissions officer at a university makes institutional decisions about hundreds of applicants with the confidence that although an institutional decision may have an adverse effect on an individual applicant, the collective good of the institution is being protected in the decision making process. In addition, group performance can be predicted better than individual performance.

Individual decisions on the other hand generally are made in a decentralized situation which is unique to the individual. Examples of such decisions are whether one should apply for a job or admission to a highly regarded university. It may be helpful to examine other people's choices and success rates; however, it still comes down to the individual's decision and the related consequences. It is the idea of individual choice and the right of the individuals to have an opportunity to make decisions about their educational and occupational futures.

Criteria for Evaluation of Tests

The classical criteria for evaluating a test are standardization, objectivity, reliability, and validity. These four categories, which are typically used to judge a test, are supplemented by a prior set of common sense considerations; namely, the purpose for which the test is to be used, the feasibility of use, the ease of administration and scoring, and finally, the cost of the test versus the available funds. Once these factors have been assessed, the more technical aspects associated with the criteria are used to evaluate the test.

Standardization means that a test has been given to a well defined population (i.e., college freshmen, auto mechanics) and that an accurate record has been kept of the scores, which are then developed into norms. The process of standardization reduces errors of interpretation about test scores. It should help reduce misunderstandings about the group with which an individual is being compared and the manner in which the individual and group performance is expressed (e.g., percentiles, standard scores).

Objectivity means free of personal bias in the scoring of the test; that is, the degree by which different people scoring the test obtain the same results. The error in scoring refers to personal errors made by the rater in scoring the test. All other things being equal, the greater the objectivity of a test the greater the reliability.

Reliability is the measure of consistency of a test or measuring instrument. Reliability relates to the extent that a test is free of variable errors when a person is measured twice by the same instrument or when one administration of a test yields small errors from person to person taking the test. A test with high reliability is one that will yield very much the same score results for a group of people under different conditions or situations.

Validity is the single most important criterion in the evaluation of a test. A test must accurately measure what it reports to measure for it to be valid. There are three types of validity: content, empirical (concurrent, predictive), and construct. The most important type of validity in aptitude testing is empirical (predictive).

The three types of validity are involved with the need to gather evidence and/or data about what the test measures. Two types of methods are generally used to gather evidence and data about the test. Logical analysis of the test-taking situation, test content or behavior and the test itself is involved in determining content validity. An empirical method (collection and analysis of data) is used to establish empirical validity. This approach is restricted to the establishment of a relationship (correlation coefficient) between test performance or behavior and a criterion performance or behavior. Construct validity relates to the gathering of evidence using both logical and empirical methods. To establish construct validity it is necessary to test hypotheses and demonstrate the theoretical test behavior is verifiable experimentally. Consequently, every validation process becomes an evaluation of the test, theory, and concept being measured. Finally, it should be kept in mind that a test is valid only for a particular predefined situation.

Questions To Be Asked about the Criteria for Evaluation of Tests

- **Standardization**
  1. How was the sample selected for the norming population?
  2. Who was included in the norming population?
  3. What are the limitations of the derived score used?
- **Objectivity**
  1. Is the scoring system free of error (e.g., same score arrived at no matter who scores the test)?
  2. What is the method of scoring?
- **Reliability**
  1. How was the reliability determined?
  2. What is the estimate of reliability for the measure? (test)
  3. What is the standard error of estimate for the test?
- **Validity**
  1. Does the test have validity for the situation in which it is being used?
  2. Does the test measure the information and/or performance on an important set of tasks?
  3. Does the test measure current performance when compared to a standard or criterion measure?
  4. Does the test measure future performance when compared to a standard or criterion measure?
  5. Does the test measure a trait or set of characteristics and can an experimental condition be created to test the hypotheses?

Criterion-Reference Measurement

NEA supports the use of criterion-reference measurement, which is generally concerned with the assessment within an individual's personal growth and, specifically, with what a person can do right now. This type of measurement is concerned with the entire educational program and not just sample items that have differentiation ability and meet other psychometric criteria.

The National Assessment of Educational Progress measurement techniques qualify as a criterion-reference approach and also incorporate many of the ideas proposed by Oscar Buros. For example, an excerpt from a National Assessment of Educational Progress report states that:

> **Methodology**
>
> To measure the nation's educational progress, National Assessment estimates the percentage of respondents (at any of four age levels) who are able to acceptably answer a question or perform a task. Each question or task (called an exercise) reflects an educational goal or objective. The exercises are administered to scientifically selected samples which take into account such variables as size and type of community, race and geographic region. Students are sampled at three age levels that represent educational milestones attained by most students: age 9, when most students have been exposed to the basic program of primary education; age 13, when most students have finished their elementary school education; and age 17, when most students are near completing their secondary education. To accurately reflect the skills, knowledge, and attitudes of the 17-year-olds, National Assessment also samples 17-year-olds not enrolled in school. Adults (ages 26 to 35) are assessed to determine the skills, knowledge and attitudes of those who have completed their formal education and have probably been away from school for a number of years. The samples are designed so that valid inferences can be made about the populations from which the samples were selected.
>
> National Assessment does not use total performance scores for individual respondents because its main concern is how various groups of individuals perform on specific exercises. Thus, it is not necessary for each respondent to take every exercise. The exercises are sorted into booklets, and each in-school respondent takes only one booklet. The various booklets are administered to statistically equivalent samples so that group comparisons can be made across booklets. This technique allows National Assessment to assess performance on far more exercises than would be possible in the usual one-hour testing situation and provides broader coverage of the assessment objectives for each learning area.
>
> Since individuals are not ranked according to their performance on the assessment materials, National Assessment does not emphasize the use of exercises with high discrimination power. The aim of the project is to describe attainment; this is best accomplished if the exercises used cover the entire spectrum of difficulty, from very easy tasks to the most difficult.

While multiple-choice exercises predominate, many open-ended exercises requiring anywhere from a few words to a long essay as an answer are included in each assessment. Exercise writers are instructed to use the exercise format that provides the best and most direct measure of the objectives being assessed. They are encouraged to develop exercises that employ the use of pictures, tapes, films or practical, everyday items as stimuli. In many assessments, individual interviews and observations of the respondents' problem-solving techniques supplement the usual data-collecting procedures. For example, in Music, respondents were asked to sing a song or perform on an instrument; in Science, respondents were asked to conduct a small experiment; in Social Studies, respondents were asked to interpret an election ballot.

National Assessment regards positive attitudes toward or opinions about the various learning areas as important educational attainments. Therefore, affective exercises and attitude survey questions are also included in most assessments.

Assessment exercises are administered either to individual or to small groups (generally less than 25) by specially trained personnel. Exercises are administered to all out-of-school respondents in a one-to-one situation. Some of the in-school respondents are assessed in a one-to-one situation. This proportion varies according to age level, learning area and assessment year. Exercises specifically designed for individual administration include those having unusual stimuli or requiring something other than a written response. In group administrations, instructions and the exercises themselves are presented to the respondents on paced tape recordings to assure complete and uniform presentation of instruction and to give those who have a reading problem a chance to hear the exercise as they are reading it.

To report the nation's educational progress, the project releases up to one-half of the exercises administered in a learning area. Released exercises allow the public to evaluate the exercises and the accompanying data. The other exercises are kept confidential and are used to assess performance changes over time.

How do NAEP test booklets differ from standardized achievement tests? Standardized achievement tests are norm-referenced; National Assessment tests are content or objective-referenced. On a standardized achievement test, each respondent takes every exercise, receives a score for his performance, and is ranked on the basis of that score with respect to a reference group. No respondent takes all of the exercises National Assessment uses to assess a learning area, no respondent receives a score, and emphasis is placed on the performance of groups of respondents on specific exercises. Standardized achievement test items are usually limited to a multiple-choice format; National Assessment employs a wide variety of exercise formats. Standardized tests usually focus on the cognitive domain; National Assessment usually includes exercises relating to the affective domain as well. A respondent is required to read the items himself when taking most standardized tests; Assessment exercises are read to respondents by a paced tape or the exercises administrator in an interview situation except during the Reading Assessment. The items on a standardized achievement test are rarely, if ever, made public; the Assessment releases up to half of the exercises used in an assessment to accompany the data.

73

NEA believes that all—not just half—of the exercises should be made public, given the fact that NAEP is funded by federal contracted dollars. Actually, the proposed FY 1980 U.S. Office of Education budget had the following testing and evaluation funds proposed.

- ESEA Title I, Sec. 183, $16.5 million, to improve state and local evaluations—assisting districts in revising their computer programs, in selecting appropriate achievement tests, and in preparing reports for local school boards.
- Bilingual Education, ESEA VII, $6 million, for studies and evaluation, including testing.
- ESEA Title IX, Sec. 922, $2 million, achievement testing assistance.
- ESEA Title IV, D, Guidance, Counseling, and Testing, $18 million, for assistance to state and local education agencies.
- Education of the Handicapped, Regional Resource Centers, $7 million, to improve the use of testing in diagnosis and prognosis.
- Women's Education Equity Act, amount not determined; however, justification for the $10 million includes funds for testing.

NEA believes that all tests or evaluation measures funded from federal dollars should be made available to the individual test taker. Moreover, no contracts should be given to any company that does not agree to provide test items, studies, or use information to the test taker. NEA strongly encourages the use of Buros' concepts and the National Assessment of Educational Progress as the models for assessment programs in education.

Finally, the NEA refers those interested in testing reform to a statement made by Robert L. Thorndike. In the first chapter of *Educational Measurement*, Thorndike examines the "state of the art" (1971) of testing. Among the topics discussed are sequential testing (tailoring testing to each specific individual being tested) and item banks and computer-generated tests. The item banks concept is very important in light of the proposed legislation (H.R. 4949). Thorndike states:

> The large memory storage available in modern computers makes it possible for testing organizations to prepare files classified by content and level in which large numbers of test items can be stored. Stored with each item can

> be all the summary statistics describing its properties of difficulty and discrimination. When a test is then desired having certain specifications of content coverage and difficulty level, an appropriate inquiry can be addressed to the computer, and it will select and display items that meet the specifications. Final assembly of the test from among the items offered from the library of items in the computer's memory storage can be made by the test editor. The time, however, may come when, for large-scale testing programs with frequent administrations of alternate forms of a test, the assembly of complete test forms will be left almost entirely to the computer. The obvious application of this technology is to large-scale testing programs with frequent test administrations, such as those of the College Entrance Examination Board and the American College Testing Program.[3]

In conclusion, NEA strongly urges the Congress to pass this vital legislation because it will help ensure that standardized tests will not deny equal educational opportunity to all people.

---

[3] Thorndike, Robert L., editor. *Educational Measurement*. Second edition. Washington, D.C.: American Council on Education, 1971. p. 6.

APPENDIXES

A. NEA's Analysis of H.R. 3564 and H.R. 4949

B. NEA's Letter of Support to the Education Commission of the States regarding the National Assessment of Educational Progress

Appendix A

Two legislative proposals concerning educational testing are before the Committee on Education and Labor. The first proposal, referred to as "Truth-in-Testing Act of 1979" (H.R. 3564), was introduced by Rep. Gibbons. The second proposal, the "Educational Testing Act of 1979" (H.R. 4949), was introduced by Rep. Weiss. The latter proposal (H.R. 4949) is based on New York legislation proposed and passed during the summer of 1979.

H.R. 3564 and H.R. 4949 concern the use of standardized tests, a subject about which NEA has raised questions and expressed concerns. Because of the NEA concern with the use of standardized tests, both proposals have been analyzed in terms of their similarities, their differences, and their responsiveness to NEA concerns.

In general, NEA believes that the two proposals represent somewhat different approaches to the use of standardized tests. To the extent that H.R. 3564 and H.R. 4949 are responsive to NEA concerns, both proposals should be supported. The Gibbons "Truth-in-Testing Act" (H.R. 3564), however, is expected to generate more opposition in Congress and could, if passed, prove to be a less successful vehicle for meeting the concerns expressed by NEA.

Both H.R. 3564 and H.R. 4949 represent notice and disclosure legislation. They differ substantially as to the type of tests covered, the extent of involvement of the Commissioner of Education and the type of enforcement provisions. H.R. 3564 would cover the National Teacher Examination which is a concern of NEA. The bill would also cover other occupational tests, which will engender opposition, and tests other than standardized tests, regulation of which would probably prove unworkable. For the most part, the disclosure requirements of H.R. 3564 require the type of information currently provided voluntarily by testing agencies such as ETS. Because H.R. 3564 does not require disclosure of underlying data on the examinations, it would not enable professionals outside the testing industry, including teachers, to analyze or comment on test construction and validity. In addition, H.R. 3564 fails to provide for disclosure of scoring data in addition to test scores which may be given to educational institutions. Groups favoring testing disclosure laws have stated that testing agencies provide information such as suspicions of cheating, unacknowledged repetition of a test and factors based on current school attended to be used in evaluating the score. Students have not been informed of this type of information where it is incorrect. In addition, students have not been provided with their test answers and the correct answers. Thus, students have been unable to learn of or correct computer grading errors. The Gibbons bill does not address this problem either.

In contrast, each one of the instances noted above is addressed in the Weiss bill with the exception of occupational testing. Various portions of the Weiss bill could use clearer and better language. In addition, some consideration should be given to the viability of including financial regulation of the testing companies in this legislation.

In addition to standardized tests, H.R. 3564 covers "oral" tests, "practical" tests and "demonstration" examinations. Sec. 2(3). The bill apparently reaches practical or demonstration examinations used in occupational licensing such as barbering, oral examinations such as the foreign service examinations, and practical or demonstration examinations used in educational admissions such as submission of a portfolio to an art school or a stage performance required for a drama school. Regulation of such tests would probably be unworkable.

H.R. 3564 contains both pre-test (Sec. 6(a)) and post-test (Sec. 6(b)) disclosure requirements which require information to be provided to test takers. Prior to administration of the test, each applicant must be provided with a written notice containing essentially the types of information currently provided voluntarily by the testing companies:

(1) a detailed description of the area of knowledge or the type of aptitude that the test attempts to analyze;

(2) in the case of a test of knowledge, a detailed description of the subjects to be tested;

(3) the margin of error or the extent of reliability of the test, determined on the basis of experimental uses of the test and, where available, actual usage;

(4) the manner in which the test results will be distributed by the testing entity to the applicant and to other persons; and

(5) a statement of the applicant's [post-test notification] rights.

The post-test notification provision requires that "promptly upon completion of scoring" the test taker must be notified of:

(1) the individual's specific performance in each of the subject or aptitude areas tested;

(2) how that specific performance ranked in relation to the other individuals and how the individual ranked on total test performance;

(3) the score required to pass the test for admission to such occupation or the score which is generally required for admission to institutions of higher education;

(4) any further information which may be obtained by the individual on request.

Section 6(c), the final substantive provision of the bill, prohibits the scoring of achievement tests on the basis of a curve:

(c) No educational or occupational admissions test which tests knowledge or achievement (rather than aptitude) shall be graded (for purposes of determining the score required to pass the test for admission) on the basis of the relative distribution of scores of other test subjects.

The enforcement provisions of H.R. 3564 (Sec. 7) authorize private causes of action by an aggrieved individual "whenever any person has administered or there are reasonable grounds to believe that any person is about to administer any ... test in violation of this act." The bill specifically provides for "preventive relief" including a permanent or temporary injunction and restraining orders and for appointment of counsel "in such circumstances as the court may deem just." The bill authorizes attorney's fees (Sec. 7(b)) and provides for federal court proceedings without regard to exhaustion of remedies. Sec. 7(c).

The enforcement procedures of injunction or restraining order represent onerous remedies, and it seems doubtful that federal courts will be inclined to enjoin the administration of standardized tests such as the SAT. For this reason, the remedies provided by the bill appear to be ineffective. Since the bill specifically authorizes "a civil action for preventive relief" courts may find that such relief is the exclusive remedy for violations of the Act.

The "Educational Testing Act of 1979" (H.R. 4949) identifies three legislative purposes (Sec. 2(b)):

(1) to ensure that test subjects and persons who use test results are fully aware of the characteristics, uses, and limitations of standardized tests in postsecondary education admissions;

(2) to make available to the public appropriate information regarding the procedures, development, and administration of standardized tests; and

(3) to protect the public interest by promoting more dependable knowledge about the limits of appropriate usage of standardized test results and by promoting greater accuracy, validity, and reliability in the development, administration, and interpretation of standardized tests.

This bill requires more extensive pre-test disclosure to test takers than H.R. 3564 and, unlike H.R. 3564, specifically requires that the pre-test notice be provided contemporaneously with the test registration form. Sec. 3(a). The legislation specifically addresses the coachability issue and requires testing agencies to inform individuals of the extent to which their scores may be improved by taking a preparation course. Pre-test notice must include the following information:

(1) The purposes for which the test is constructed and is intended to be used.

(2) The subject matters included on such test and the knowledge and skills which the test purports to measure.

(3) Statements designed to provide information for interpreting the test results, including explanations of the test, and the correlation between test scores and future success in schools and, in the case of tests used for post baccalaureate admissions, the correlation between test scores and success in the career for which admission is sought.

(4) Statements concerning the effects on and uses of test scores, including --

  (A) if the test score is used by itself or with other information to predict future grade point average, the extent, expressed as a percentage, to which the use of this test score improves the accuracy of predicting future grade point average, over and above all other information used; and

  (B) a comparison of the average score and percentiles of test subjects by major income groups; and

  (C) the extent, if available to the test agency, to which test preparation courses improve test subjects' scores on average, expressed as a percentage.

(5) A description of the form in which test scores will be reported, whether the raw test scores will be altered in any way before being reported to the test subject,

and the manner, if any, the test agency will use the test score (in raw or transformed form) by itself or together with any other information about the test subject to predict in any way the subject's future academic performance for any postsecondary educational institution.

(6) A complete description of any promises or covenants that the test agency makes to the test subject with regard to accuracy of scoring, timely forwarding or score reporting, and privacy of information (including test scores and other information*), relating to the test subjects.

(7) The property rights of the test subject to the test results, if any, the duration for which such results will be retained by the test agency, and policies regarding storage, disposal, and future use of test scores.

(8) The date by which the test subject's test score will be completed and mailed to the test subject.

(9) A description of special services to accommodate physically handicapped test subjects.

---

* The phrase "and other information" was added by Weiss' staff subsequent to conversations with NEA. Significant questions exist as to the use made by ETS of personal data obtained on the test or test application. ETS sells student lists to institutions.

In addition to providing notice to test subjects, the bill requires the testing agency to provide the same information to the recipient institution prior to or coincident with the reporting of test scores.

The major area covered by the Weiss bill is reporting to governmental educational agencies. Two types of information must be disclosed to the government. First, this reporting requirement concerns the studies and evaluations of the tests themselves and is designed to allow professionals outside the testing industry, including teachers, access to such studies to allow independent analysis of the construction, validity and use of the tests. The second type of information to be disclosed includes the test questions and answers and scoring rules. This is accomplished by cross-reference to the Freedom of Information Act, 5 U.S.C. § 552(a)(3), which authorizes release of records. The test agency is required to provide to the Commissioner of Education:

> (A) Any study, evaluation, or statistical report pertaining to a test, which a test agency prepares or causes to be prepared, or for which it provides data. Nothing in this paragraph shall require submission of any reports or documents containing information identifiable with any individual test subject. Such information shall be deleted or obliterated prior to submission to the Commissioner, [and]
>
> (1) shall, within 30 days after the results of any standardized test are released, file or cause to be filed in the office of the Commissioner --

(a) a copy of all test questions used in calculating the test subject's raw score;

(b) the corresponding acceptable answers to those questions; and

(c) all rules for transferring raw scores into those scores reported to the test subject and postsecondary educational institutions together with an explanation of such rules.

This data, in addition to being made available pursuant to the Freedom of Information Act, must be made available by the Commissioner of Education to state educational agencies and commissions.

The testing agency must also provide the questions, the correct answers, and the test taker's answers, as well as scoring information, to the test subject on request for a ninety-day period subsequent to release of the test scores.

Furthermore, the legislation requires the Commissioner of Education to prepare for Congress an evaluation of the data on these tests both with regard to coachability and cultural bias:

> (b) The Commissioners [sic] shall report to Congress within one year of the effective date of this Act concerning the relationship between the test scores of test subjects and income, race, sex, ethnic, and handicapped status. Such report shall include an evaluation of available data concerning the relation-

ship between test scores and the completion of test preparation courses.

The major difference between the Weiss draft and the New York law upon which it is based is an attempt in the federal legislation to regulate the costs to test subjects of the tests and to require financial disclosures by the testing companies. During the New York hearings, the testing companies argued that costs would skyrocket Proponents of the New York legislation, New York Public Interest Research Group and Nader in particular, questioned these predictions using whatever data they could obtain from the testing companies, especially ETS. Section 7 of the bill entitled "Testing Costs and Fees to Students" provides as follows:

> In order to ensure that tests are being offered at a reasonable cost to test subjects, each test agency shall report the following information to the Commissioner:
>
> (1) Before March 31, 1981, or within 90 days after it first becomes a test agency, whichever is later, the test agency shall report the closing date of its testing year. Each test agency shall report any change in the closing data of its testing year within 90 days after the change is made.
>
> (2) For each test program, within 120 days after the close of the testing year the test agency shall report--
>
> (A) the total number of times the test was taken during the testing year;
>
> (B) the number of test subjects who have taken the test once, who have taken it twice, and who have taken it more than twice during the testing year;
>
> (C) the number of refunds given to individuals who have registered for, but did not take, the test;

(D) the number of test subjects for whom the test fee was waived or reduced;

(E) the total amount of fees received from the test subjects by the test agency for each test program for that test year;

(F) the total amount of revenue received from each test program; and

(G) the expenses to the test agency of the tests, including --

- (i) expenses incurred by the test agency for each test program;
- (ii) expenses incurred for test development by the test agency for each test program; and
- (iii) all expenses which are fixed or can be regarded as overhead expenses and not associated with any test program or with test development;

(3) If a separate fee is charged test subjects for admissions data assembly services or score reporting services, within 120 days after the close of the testing year, the test agency shall report --

(A) the number of individuals registering for each admissions data assembly service during the testing year;

(B) the number of individuals registering for each score reporting service during the testing year;

(C) the total amount of revenue received from the individuals by the test agency for each admissions data assembly service or score reporting service during the testing year; and

(D) the expenses to the test agency for each admissions data assembly service or score reporting service during the testing year.

The Weiss bill, like the New York legislation, uses a civil penalty as its remedy. While the New York law establishes a $500 penalty per violation, the federal law establishes a $2,000 fine. This would represent a small penalty where the test agency failed to properly report to the Commission of Education since this would probably constitute a single violation. With regard to violations of the notice to students the penalties could be substantial since presumably failure to provide the required notices to students would result in multiple violations reflecting the number of students involved. One potential difficulty in enforcement may be in determining which and how many individuals were not given proper notice or timely reporting. The Commissioner is authorized by the draft to promulgate regulations to implement the legislation and enforcement would be one area where regulations might fill in the sketch created by the draft legislation.

The Weiss bill would require disclosure to students of covenants and promises made by the testing agencies. Private causes of action by test takers could be based on breaches of these contractual warranties.

Appendix B

EXECUTIVE OFFICE

nea

NATIONAL EDUCATION ASSOCIATION • 1201 16th St., N.W., Washington, D C 20036 • (202) 833-4000
JOHN RYOR, President
WILLARD H. MCGUIRE, Vice-President
JOHN T. MCGARIGAL, Secretary-Treasurer
TERRY HERNDON, Executive Director

June 8, 1979

Dr. Warren G. Hill
Executive Director
Education Commission of the States
1860 Lincoln Street
Denver, Colorado 80203

Dear Dr. Hill:

The National Education Association strongly supports the Education Commission of the States' application to continue as the organization responsible for the National Assessment of Educational Progress. The National Assessment has gained respect from teachers, administrators, and educational policy makers at all levels of the education community over the last fifteen years.

NEA advocates measurement techniques and approaches which help policy makers formulate intelligent decisions about school programs. The National Assessment has provided this information in the past and it is hoped that the program can be extended down into the local school districts to replace the current fad of competency testing.

NEA strongly supports the makeup of the National Assessment Policy Committee which includes teacher representation on the committee. The Association would urge that the Federal Government continue this practice and require that teachers be represented on the National Assessment Policy Committee in direct proportion to their national membership. The four teachers on the committee should be designated by the majority organization or, if this is not possible, allocated to teacher organizations according to membership. Administrator and school board organizations should designate their representatives.

The NEA recommends that the ECS be granted the funds to continue the NAEP.

Sincerely,

Terry Herndon

Terry Herndon
Executive Director

Chairman **Perkins**. Mr. Gould, the main point of controversy regarding the Weiss bill is over the requirement for disclosure of the tests themselves and of all of the corrected answer sheets.

Could you tell us again precisely why the General Accounting Office opposes the imposition of this requirement on the Federal Government?

Mr. **Gould**. One of the primary reasons is that there would be a substantial economic burden if the Federal Government were required to provide copies of test questions and test answers to all examinees. Such procedure would also preclude the reuse of any of those test items.

New ones would have to be prepared each time a test is given, and this has technical implications. The cost implications are illustrated in an example prepared by OPM. The government's clerical exam is administered nationwide every working day. If test items were disclosed, 250 separate and unique exam booklets would be needed each year.

It takes about one staff year, at an average salary of $20,000, to prepare a clerical exam booklet using existing questions. This does not include additional administrative costs required to insure that the same test is given nationwide on each day so that applicants in one area do not benefit from disclosure in another area.

The administrative problems, budget, and staff requirements are multiplied many times when more complicated, higher level exams are used. To develop the current pool of items for the PACE, for instance, required a total of 52 staff-years of effort over a five-year period, and the cost and workload estimates I have just provided are for test development only. They do not cover the very time-consuming and costly process of validating the test once it is developed.

From the technical standpoint there is finite limit to the number of new questions that can be developed in any subject matter area without quality deteriorating to a level that the exam would no longer be an effective means of evaluating people for employment.

In the second place, a group of common questions is needed in each version of a test so that adjustments can be made for differences among the versions in average difficulty. Only in this way can a score have the same meaning on one version as it has on another. Disclosure would make it impossible to have a common group of questions.

Chairman **Perkins**. Now, what in your judgment would be the additional cost required under the Weiss bill, assuming that it were adopted?

Mr. **Gould**. There would have to be a test prepared each time those questions and answers were provided to the test applicants, and we are talking about a test that 135,000 individuals took in 1978, of which 77,000 passed with scores of 70 or above; and 7,600 were hired. This test is given four times a year, so you would need four separate tests. This test took an average of 5 years just to develop the test and there has been extensive research going on to show the validity of this test which is also a very costly procedure.

Each of these tests would have to be shown to be job-related to be able to use it. If you didn't have good validation you would end up with many of these tests in the courts trying to determine whether they were or were not job-related.

Chairman PERKINS. Sticking to the Weiss bill, could you give us a figure of the additional cost in your judgment that would be involved if it were adopted?

Mr. GOULD. Perhaps we ought to consult with the OPM people that provided the validation. It took $2 million to develop the PACE test, not including the validation costs, so it would be a substantial amount of money for each test.

Chairman PERKINS. Mr. Erdahl?

Mr. ERDAHL. Mr. Gould, I have before me here a report from the Comptroller General, a report to the Congress involving testing that came out on May 15, 1979.

Are other studies like this being prepared at the present time, or has there been one subsequent to this?

Mr. GOULD. No, sir; that is the only study that we have made in that area. We have recommended that OPM collect data on the race and sex of the applicants taking the tests so that they can come up with comparable kinds of information in the future.

You may note from that particular report that it was very difficult for us to get information on the race and sex of individuals. We had to do it by taking social security numbers of the individuals who sat for the examination, asking Social Security Administration to identify in gross numbers for us the race and sex of those individuals.

It is a very difficult and time consuming review which took in the neighborhood of about 2 years of our time to do.

Mr. ERDAHL. You said there is another study that involves more about race and sex, how people do in the test. Is that what it applied to?

Mr. GOULD. OPM now is collecting that kind of information on a routine basis so they can provide you that kind of information on current tests.

Mr. ERDAHL. Your main concern, to echo, I guess, the question that the chairman asked and you responded to, is that really it would be a cost burden. I have the assumption from what your statement said that you still were going to have to do some type of a selection process to try to pick the qualified employees for the Federal Government.

Mr. GOULD. That is correct, sir. We are very concerned that the test itself does screen out a large proportion of minorities compared with nonminorities.

We have asked OPM to look at alternative procedures that are equally valid in terms of selecting qualified people. We do not know of any alternative at this point in time which would allow you to screen the large number of individuals that have to be screened for Federal employment.

We believe that some sort of screening is necessary, at least until a suitable alternative is found.

Mr. ERDAHL. Maybe you were at some of the other hearings on this subject, but several of us talked about the concern over this so-called truth in testing, because certain groups of people do poorly in the tests.

We made the comparison, we are shooting the messenger because he brings bad news. What really concerns us is that there are some flaws in the educational system itself that people are not receiving the proper educational opportunities to do well in the test; and if that is the case, we are putting our emphasis in the wrong area and taking after the test, when maybe we should be taking after the system.

Mr. Chairman, I have no further questions at this time.

Mr. Weiss. Let me apologize for arriving late. I am not in the best of shape, as you can tell from my voice. That may be of benefit to everyone, because I will limit my questions.

I understand, Mr. Gould, in your prepared testimony you had indicated that there would be a problem of costs in the development of new test questions if H.R. 4949 were to be adopted; is that correct?

Mr. Gould. Not only in terms of costs but the feasibility of coming up with continually new items for the test. There is not an unlimited supply of questions that would be valid. We also pointed out that with the new test you would still have the extensive validation process that you have to go through to show job-relatedness, so it's more than just the cost aspect.

Mr. Weiss. Perhaps you testified to this in your earlier testimony; would you give me some background as to your credentials in the testing field?

Mr. Gould. I personally have none, sir. Dr. John Harper is the personnel psychologist who spent quite a bit of time in that area, and let me give you his qualifications, if I may.

Mr. Weiss. You personally are not testifying on the basis of your own knowledge or expertise, but on the basis of Dr. Cameron's?

Mr. Gould. No; on the basis of Dr. Harper and the other experts within the General Accounting Office. I am not the expert in the General Accounting Office. Dr. Harper who sits on my right here is.

Mr. Weiss. Perhaps we ought to get his credentials in the record.

Dr. Harper. I have a doctorate in industrial and organizational psychology. Is this the type of information you are looking for?

Mr. Weiss. Sure.

Dr. Harper. I received my masters degree and did training under Ted Cureton, who was one of the leaders in the field of psychometrics, one of the early leaders in that field.

I have been working in the General Accounting Office for the last 3 or 4 years in the area of survey research and evaluation research, and particularly looking at the evidence on the validity of the PACE developed by the Civil Service Commission and how that fits with professional practices, and also with the uniform guideline requirements of the Federal Government.

Mr. Weiss. Have you had occasion to follow any of the testimony that we have had in the course of the series of hearings on H.R. 4949?

Dr. Harper. I am not familiar with all of the testimony that you have been developing on these two bills, but I am familiar with some of the testimony.

Mr. Weiss. You are familiar with testimony from experts who testified here; that is, people who are involved in the field of testing although not associated with the testing industry as such, who have testified as to the problems that they saw in the tests themselves, as far as the inherent, although probably unintentional biases that have crept in on the basis of geography, sex, and so on?

Are you familiar with that aspect of the testimony?

Dr. **Harper**. Not particularly with respect to the testimony you have been taking during those hearings, but certainly with respect to what has been appearing in both the professional and public literature in this area.

Mr. **Weiss**. Do you generally share that sentiment that indeed there is unintentional bias in those tests?

Dr. **Harper**. As we testified before Congresswoman Schroeder's subcommittee, the question of bias in testing is a very complicated question; and it requires, I think, a definition of bias on the part of those who are discussing this issue, so that if you were to talk about bias in terms of whether or not a test is related to job performance or academic achievement, then the conclusion with respect to what is appearing in the professional literature is clearly tests are not biased.

If you were to talk about bias in terms of how different groups of people perform on the tests, if that is your definition of bias; that is, that some groups of people score higher on tests than other groups of people do, then tests for the most part are biased.

Those are the two most popular definitions. There are others.

Mr. **Weiss**. Using that last definition, that is, of different groups of people doing differently on the examinations, again, just so we agree as to what our ground rules are as to what we are talking about, not on the basis of some people scoring higher at random or lower at random, but entire groups on the basis of sex, or geography or race. Is that what you mean by saying that different groups of people score differently?

Dr. **Harper**. Yes; that is the second definition of bias that I would use, and I would just mention, to my knowledge, there are very little sex differences in test performance.

Males and females generally perform about the same generally, although there are some differences.

Mr. **Weiss**. Do you agree as to that second definition of bias? There is validity to the complaint that probably unintentional but nonetheless there is in fact bias in those tests?

Dr. **Harper**. I personally and professionally wouldn't because I would ask the next question: That is, are these differences in test score performance reflected in performance on the job or in the educational achievement?

Based on the answer to that next question, I would make my decision about whether the tests are biased.

Mr. **Weiss**. Again, if you haven't heard the testimony, it's very difficult or unfair for me to press you on it.

If you reread some of the testimony, read some of the testimony we have had, you will in fact note that there has been testimony before us from people who are expert in the field of testing who have testified very clearly and unequivocally that there is bias in the test basis, geography, on the basis of race, sex, and so forth.

Dr. **Harper**. OK; that is using the second definition.

Mr. Weiss. Using the second definition. Let me ask you as to the testimony which Mr. Gould gave on the basis of your advice and on the advice of your experts with GAO.

On what basis do you come to the conclusion that there is a finite number of questions that could be asked in any of those various areas that are tested by the standardized tests; and, two, on what basis did you come to the conclusion as to the additional expense involved in developing additional questions, if in fact H.R. 4949 were to be passed?

Dr. Harper. On our first question relating to the finiteness of the universe of the test questions, I think if you consider for a moment particularly the trades test area, it would appear intuitively reasonable that after some point you cannot generate anything other than different ways of asking the same question; that is, you can only generate new ways of asking what is essentially the same question.

Mr. Weiss. Assuming for our purpose that is true, although not accepting it as true, H.R. 4949 does not in any way address itself to trade professional testing. It simply addresses itself to admissions to institutions of postsecondary education.

Mr. Gould. Our testimony was directed at extending those two bills to Federal testing which would encompass those two areas that you said were excluded from the legislation.

As it is now written, it would have no impact on the Federal sector, but we were asked to testify on the impact it might have on the Federal sector.

Mr. Weiss. Your testimony in no way addresses itself to or applies to H.R. 4949 as presently written?

Mr. Gould. Not as presently written, only as they would be extended to the Federal sector.

Mr. Weiss. Thank you very much.

I have no further questions on that basis.

Mr. Erdahl. If I might respond to your line of questioning to the witnesses here, it seems to me there is a certain relevance to it, because if we say this applies to people that are seeking entrance to higher education, but that the same principles shouldn't apply to the many thousands of people that apply, and those that are hired by the Government, there is a definite correlation, because the tests are obviously used as an indication of the person's ability to perform in an academic or occupational situation, so I think that their testimony is relevant.

If there is no objection, Mr. Chairman, I would like to have inserted into the record an article that appeared in the New York Times on October 8 of this month, entitled "Test Publishers See Dropping of Exams," and the second one is "Most Entrance Tests May End in State."

We understand that this bill is patterned after the New York law, and author of that law, Senator LaValle, acknowledged, that the economic cost of special administrations type of tests posed legitimate problems. He said he planned to convene the State Higher Education Committee to discuss possible changes in the legislation.

Such changes could be made when the legislature convenes probably in November to set a date for primary elections. It seems to me that even in New York where this concept is being implemented there are already questions about it; so I hope that we, on this committee, would also be willing to investigate those questions.

Mr. **Weiss**. Without objection, that will be entered into the record.

[The newspaper article referred to follows:]

[From the New York Times, Oct. 8, 1979]

TEST PUBLISHERS SEE DROPPING OF EXAMS

SAY NEW YORK'S "TRUTH" LAW WILL CUT MOST IN ADMISSIONS USE

(By Edward B. Fiske)

Three-quarters of the standardized college and graduate school admission tests covered by the so-called "truth in testing" law in New York State will probably not be given in the state after the law goes into effect Jan. 1, according to the State Education Department.

Sponsors and publishers of 20 of the 26 tests covered by the law have concluded that it is economically impossible to comply with the requirement to publish questions within 30 days of reporting students' scores, according to David R. Bower, administrator of the office of post-secondary assessment and testing of the Education Department.

"The law is having the greatest impact on tests with low volumes," he said. Students will still be able to take the Scholastic Aptitude Tests and other major examinations that are given to large numbers of students.

UNABLE TO REUSE OLD "FORMS"

The tests expected to be withdrawn include those routinely used for admission to professional schools of medicine, dentistry, pharmacy, optometry and nursing, and the Miller Analogies Test, which is used in a variety of graduate fields. Citing the disclosure requirement, publishers and sponsors of the tests say they will not be able to reuse old forms of their tests.

With regulations for implementing the law still in the drafting stage, officials of the State Education Department and others are still unclear about the ultimate impact of the new law on both individuals and institutions.

"I suppose some schools will have to simply drop the use of test scores and rely on other kinds of information," Mr. Bower said, "But then colleges say that they never rely only on test scores in making their decisions anyway. For students with good grades, this will probably be all right, but for others it could cause serious problems."

Some test makers expect some students simply to cross state lines to take the tests. The State Education Department, however, has taken the view that the law applies to any test whose results are sent to an institution in New York State. If this interpretation holds, crossing state lines would be of no benefit to students planning to attend institutions in New York State. New York is the only state to have such a disclosure law.

"As we assess the situation today, we don't see how we can continue in the business in New York State" said Roger Lennon, associate to the chairman of Harcourt Brace Jovanovich Inc., owner of the Psychological Corporation, which is planning to withdraw a dozen examinations.

"We can continue to provide the programs and services in the other 49 states, and not New York, or get out of the business entirely."

Although no formal policy has yet been announced, the College Entrance Examination Board, which sponsors the Scholastic Aptitude Test and the Graduate Record Examination, is expected to continue operating in New York on a reduced level. The American College Testing Service, the other large college admission testing agency, and the Law School Admissions Council, which gives the Law School Aptitude Test, have indicated that they will follow similar policies.

"SOME LONG-TERM PROBLEMS"

The Graduate Management Admission Council, which sponsors the Management Admissions Test, has announced that it will continue to operate in the state. "The law poses some long-term technical problems, but we hope to continue our present level of service," said William Broesamle, the council chairman.

Sponsors and publishers of the tests said in interviews that the new law constituted an enormous—and in some cases prohibitive—burden.

"Each new form of the Entrance Examination for Schools of Nursing costs $25,000," said Mr. Lennon of Harcourt Brace Jovanovich, who noted that most of the tests being withdrawn by the Psychological Corporation were given to no more than several hundred New Yorkers annually. "If we were to continue giving it 19 times in New York to 1,500 candidates, the cost would be $475,000, or $317 per student."

The principal sponsor of the "truth in testing" bill, Senator Kenneth P. LaValle, Republican of Port Jefferson, L.I., said the decisions to withdraw from the state were "hasty" and premature.

"They're not even giving the law a chance," he said.

PROBLEM ON SPECIAL EXAMS

George Hanford, the new president of the College Board, said that with 250,000 New Yorkers taking S.A.T, examinations each year the cost of additional forms for the seven regular administrations each year could probably be absorbed with a modest surcharge above the regular $8.25 fee for New York students. But he added that the board faced a "major problem" with special examinations.

"We give examinations on Sundays for students who for religious reasons cannot take them on Saturdays, and we make arrangements with handicapped students to administer them virtually any day of the year," he said. "There's no way we can continue this level of service if we have to reveal the questions every time we use a new form."

The College Board is expected to announce its plans shortly. It has already indicated, however, that it will probably withdraw the Pruebas de Aptitud Academica, the Spanish version of the S.A.T., which is produced in Puerto Rico and used on a limited basis in New York.

The Admissions Testing Law of 1979 was passed by the Legislature and signed by Governor Carey in July in response to vigorous lobbying by student and consumer groups. These argued that the testing industry operated in too much secrecy and that students had a right to see and understand the examinations that pay an important part in their academic and professional careers.

ARGUMENT OF OPPONENTS

The legislation was opposed by Gordon M. Ambach, the State Commissioner of Education, and by the testing industry. They contended that the additional costs would far outweigh any education or other benefits and that there were other ways to bring about more "accountability" on the part of test makers.

The law requires manufacturers of college and graduate school admission tests to publish test questions within 30 days of reporting scores and, upon request, to give students copies of their answers. It also requires test publishers to supply the Commissioner with reports of research done in conjunction with these tests and to take certain steps to assure confidentiality.

In recent weeks, sponsors and publishers of the tests covered by the law have held a series of meetings with Senator LaValle to explore problems posed by the legislation. Some have asked for a six-month delay in the implementation date, and others have asked to have tests exempted.

Sponsors of the New Medical College Admissions Test, for example, argue that the specialized nature of their subject matter limits the number of questions—and thus new forms—that they could come up with.

"UNIMPRESSED" BY ARGUMENTS

James W. Graham, associate secretary of the Council on Dental Education of the American Dental Association said that there was a similar problem with the "perceptual ability" section of the Dental Admission Testing Program. "There are only so many ways you can draw pictures of boxes," he said.

Senator LaValle said that he was "unimpressed" with the arguments for a delay. "I get the impression that some of them just want out from the law," he said. "When you pass a major piece of legislation that changes things drastically, there is no way you can avoid a little chaos in the first six months."

He acknowledged, however, that the economic cost of "special administrations" posed a legitimate problem and said he planned to convene the Senate Higher Education Committee to discuss possible changes in the legislation. Such changes could be made when the Legislature convense, probably in November, to set a date for primary elections.

There is already a precedent for excluding low-volume tests because the law specifically exempts the achievement-test sections of the S.A.T. and the G.R.E. examinations. "There is too small a population taking these," said Senator LaValle. "There was input from Educational Testing Service, and it make sense." The testing service develops the tests for the College Board.

POLITICAL MISCALCULATION

Asked why other tests, such as the Aptitude Test for Allied Health Programs, which is given to only 200 New Yorkers every year, were not excluded on the same ground, Senator LaValle replied: "They never came and told us about the problem. They just assumed that the bill would never pass."

Mr. Lennon of Harcourt Brace Jovanovich conceded that, in not testifying at hearings on the bill, he and others at the Psychological Corporation had made a political miscalculation. "Had I felt it was more likely to pass, I might have urged us to be more active," he said.

A major issue is whether the new state law is, for all practical purposes, a "national" law.

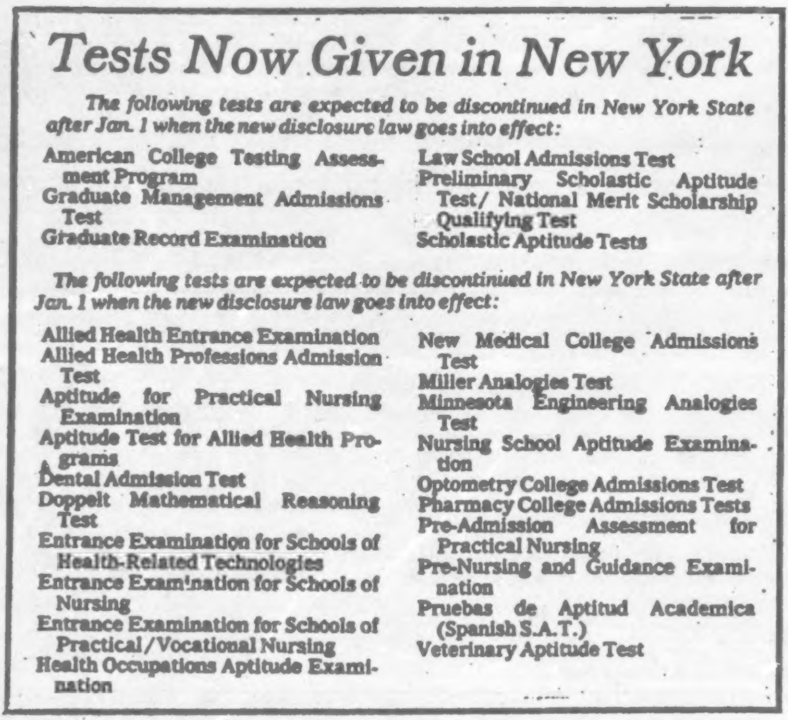
Mr. Lennon said that, in accordance with the interpretation of the New York State Education Department, the Psychological Corporation would not only stop administering tests in New York, but would also institutions and individual counselors using the Miller Analogies and two other locally administered tests not to forward the results to any New York colleges and universities.

The Law School Admissions Council indicated that it would probably adjust its entire Law School Admissions Test Program to the requirements of the New York law.

"The principle of uniformity is important," said L. Orin Slagle, president of the council. "We want to keep conditions for everyone as comparable as possible."

Other test sponsors, however, argue that the State Education Department's interpretation is not warranted by the language of the law and that, even if it is, it is unenforceable.

Whether the law is, in fact, a national one has important consequences for the cost of future tests to New York residents.



Tests Now Given in New York

The following tests are expected to be discontinued in New York State after Jan. 1 when the new disclosure law goes into effect:

- American College Testing Assessment Program
- Graduate Management Admissions Test
- Graduate Record Examination
- Law School Admissions Test
- Preliminary Scholastic Aptitude Test/ National Merit Scholarship Qualifying Test
- Scholastic Aptitude Tests

The following tests are expected to be discontinued in New York State after Jan. 1 when the new disclosure law goes into effect:

- Allied Health Entrance Examination
- Allied Health Professions Admission Test
- Aptitude for Practical Nursing Examination
- Aptitude Test for Allied Health Programs
- Dental Admission Test
- Doppelt Mathematical Reasoning Test
- Entrance Examination for Schools of Health-Related Technologies
- Entrance Examination for Schools of Nursing
- Entrance Examination for Schools of Practical/Vocational Nursing
- Health Occupations Aptitude Examination
- New Medical College Admissions Test
- Miller Analogies Test
- Minnesota Engineering Analogies Test
- Nursing School Aptitude Examination
- Optometry College Admissions Test
- Pharmacy College Admissions Tests
- Pre-Admission Assessment for Practical Nursing
- Pre-Nursing and Guidance Examination
- Pruebas de Aptitud Academica (Spanish S.A.T.)
- Veterinary Aptitude Test

Mr. **Weiss**. Mr. Erdahl, you might also be interested in having entered into the record a series of questions and answers that appears in relation to testing legislation; and it appears in this morning's front page of the New York Times and carries over and, again, primarily addresses the New York law, although it mentions in passing that Congress is considering it.

Mr. **Erdahl**. I would hope that that could be part of the record, too.

Mr. **Weiss**. It should also be noted by way of clarification that many of the problems that are inherent in the New York legislation primarily because we had the benefit of the work product that they had completed would be cured by H.R. 4949 and would never come up as a problem, because we automatically exclude those tests which are given to fewer than 5,000 people during the course of the year.

The tests that they are talking about are those which are small volume tests based by the small manufacturers, and again many of those are tests which apply beyond just admissions to institutions of postsecondary education, so I think that some of the problems that they are now discovering in the New York legislation have already been anticipated or seen by us and corrected.

I have some questions for the record which I would like to ask for the record.

Mr. Cameron, in order to maintain a balance at the hearing, and also because it's nice to hear support for the legislation from that side of the table as well as from this side of the table, would you just in brief comment on what in fact what Mr. Erdahl had raised about a story appearing in the New York Times about the conduct or behavior, or I don't know how you would characterize it as to what the testing companies, as to what their role has to be if we adopt legislation of this kind?

Do you have any view on that at all?

Mr. CAMERON. I don't. Mr. Kovacs, our director of research, might.

Mr. KOVACS. I think it's important to make a distinction between the kinds of tests that H.R. 4949 addresses and the material standardized tests which go all over the field to include tests of mechanical ability or tests of typing skill or what have you.

H.R. 4949 addresses aptitude type tests, and those tests are designed to predict future success, generally the first semester in college. The other side of the question is not how well they are doing but what they do not measure. Maybe the system is wrong, and these youngsters aren't doing well on those tests because the system isn't producing quality students.

The truth of the matter is about 75 percent of what works, motivation, desire, interest in succeeding is what makes the difference in terms of higher education or law school graduate, and we all know that once they get out into the work field that we may have another different criterion for predicting success.

None of these characteristics are really addressed by the so-called aptitude tests. They principally focus on the first semester in school and help admissions officers sort out and include or exclude based on some criteria.

Most admissions officers use more than just the test, however. The test is used as part of the decisionmaking process, and depending upon the institutions, it may make a difference on whether or not the youngster is admitted or not.

I don't know if that directly relates to your question, but I will be happy to try and clarify it.

Mr. WEISS. Again, for either Dr. Harper or you, anybody at the panel, one of the areas that we had inquired into when we had some of the testing companies here before us was the additional burden thrown upon them in the preparation of additional test questions, and also the cost implicit, and again you might want to doublecheck for yourselves the testimony that was given to us by the American College Testing Service representative who opposed the legislation, for the record—I want to make that clear—stated to us that in their instance four out of five of the test portions of any particular test is new material on every single test that they administer, and that, therefore, overall 80 percent of the test questions on any form and generally across the board are new questions.

He said in totally unequivocal terms that cost is absolutely no factor, as far as they are concerned, since they right now develop so many new questions. In any event, that the additional burden would be so miniscule, so slight, that it would not impact upon them from a cost point of view.

I make that comment only because I understand that Mr. Gould and Dr. Harper were really referring to an extension of H.R. 4949 beyond its present scope; but since there has been a great deal of cost argument in this whole discussion, indeed today's New York Times' story quotes Educational Testing Service as saying that they might have to increase the cost of their tests, the $8.25 for the scholastic aptitude test by as much as $5 per applicant.

That kind of conclusion has to go really beyond the actual cost of test development, and I think has another motivation which is to scare people into opposing our efforts to take the mystery and mystification out of the test.

I want you to bear that in mind. I have to go answer a quorum call at this point, and I will be back in 3 or 5 minutes, and then if you would like to comment on that, you may.

[A brief recess was taken.]

Mr. WEISS. The committee will come to order again.

If anyone can remember the question that I posed and if anyone wants to respond to it, now would be the time.

Dr. HARPER. I will start.

Mr. WEISS. Dr. Harper.

Dr. HARPER. I don't think it would be appropriate for us to comment on what the representative from the American college testing program said here without examining the full record of what he had to say, the context he was saying it in, et cetera.

But I think in terms of what we have been asked to address, that is, extension to Federal testing and particularly the PACE, I would say that clearly the writing of new items is a difficult problem.

We have been told by representatives at the office of personnel management how long it took to put that series of exams together, the written test for the PACE, and I have talked to some of the people that were involved in the item writing.

The item writing in some parts of that examination was a very excruciating process because the professional requirements for a good item in terms of what kind of differentiation it gives you in relation to all the other items are quite stringent.

So the process of writing high level items such as appear in PACE is a long and protracted exercise.

Mr. WEISS. Does anyone else want to comment on it?

Let me ask you in passing, didn't GAO do an audit of the civil service examinations and come to the conclusion that you were not altogether happy with the makeup of those examinations?

Mr. GOULD. Yes. We issued a report on May 15 of this last year. If you would like—oh, I see you have it there. There are a number of questions that have been raised primarily toward the validity of that test and the job relatedness of it. We felt that for the test to be in compliance with the uniform guidelines that there were some additional validation that the OPM could do to improve that examination.

We have made recommendations on some things that we thought they ought to do to help improve the validity. We also suggested and recommended to them that they look for some other alternatives, selection devices, because of the discriminatory effect that the exam had on particularly blacks who were screened out at a much higher proportion than other folks who took the examination.

We have a very perplexing problem there. This test is probably the most validated test in the history of the Federal Government. It has been shown by the researchers to be job related. Yet, blacks do not do as well on that examination as whites do. It is a very difficult problem.

We feel if there are other alternatives which will give us a better representation of minorities in the work force, that we ought to do everything we can to try to find what those alternative selection devices might be.

Mr. **Weiss**. I appreciate that. That was my recollection.

Now, staff has given me the report that you have just referred to. I find that on the very front cover of the report, the second paragraph states that GAO found that PACE and the Junior Federal Assistant Examination screen out black applicants at a much higher rate than whites and that few blacks score high enough for a realistic job opportunity.

In a sense, I assume that on the basis of your comment just now that you are not suggesting that the bias was deliberate but unintentional. There is something the matter obviously with what those tests are testing at this point.

Given that kind of conclusion on the PACE test, I find, Dr. Harper, your daintiness in approaching the question of discrimination on the non-PACE test, in the scholastic aptitude test, for example, sort of unexplainable to me.

It seems to me that your own people have found the same kind of questions and problems in relation to the PACE test that the critics of the SAT's and the other standardized tests have found in relation to it.

All that we are suggesting, and I assume that you are not going to take the position that simply because it is going to cost more to develop new tests or that it will be a burden on the people that prepared the old tests that in fact the current situation be allowed to continue.

I assume that you are going to be recommending a total review, new tests to be developed, and regardless of what the current validation seems to be as far as the in-house people are concerned, that you are not going to accept the continuation of the situation that prevails in the Federal examination.

Now in essence, without telling people what kind of questions they ought to be asking or how their questions ought to be posed, all that H.R. 4949 says is to open the process up so that experts in and out of the field, in addition to the students themselves, can tell where they were marked wrong and what in fact is going on with those tests so that in fact we can develop more accurate tests, the people in the field can develop more accurate tests.

I don't see why that should be objectionable or that should be subject to the argument that, yes, maybe it would do that, but because it would cost something additional by way of moneys it is something not worth doing.

You are not suggesting that it is not worth doing, are you?

Dr. **Harper**. Let me just say in terms of my definition of discrimination and the daintiness with which you say I walk around that, I believe that my interpretation is consistent with title VII of the 1964 Civil Rights Act and with a long list of Supreme Court decisions going back to *Griggs* v. *Duke Power* and *Albemarle* v. *Moody* and *Washington* v. *Davis* and the more recent decisions.

I recognize this as a serious problem. I share the concern Mr. Gould has expressed and the concern we put forward in this report about what is to us a perplexing problem that has not an easy solution, and one. I think hearings such as you are having today and other committees have had will hopefully shed some light on this concern and provide all of us with a better understanding of issues involved in the area of testing, ability measurement, performance both in school and on the job.

For that reason we think that these types of hearings are very worthwhile.

Mr. **Weiss**. Were you involved in the study which culminated in the GAO report of May 1979?

Dr. **Harper**. Yes, I was involved in that study and also a study which preceded this one by about 6 or 8 months which was GAO's analysis of the process, procedures and activities of the Equal Employment Opportunity Coordination Council created by the 1972 amendments to the Civil Rights Act which produced the Uniform Guidelines on Employee Selection Procedures.

So I was involved in writing that earlier report which provided the context for evaluating PACE in terms of what the Government's requirements are for testing as well as in writing the PACE report.

Mr. **Weiss**. Do you agree with the conclusions of the report as set forth on the front of the May 1979 report?

Dr. **Harper**. Yes. The facts are straightforward. There is very little argument. There is some argument over what the implications of the facts are.

I think the questions that are raised are questions about what are the conclusions based on the facts and what are the courses of actions that might follow.

I think if we go back to the earlier interchange that we had about the definition of discrimination, so long as you can say that of whenever there are group differences in performance on a test there is automatically discrimination, when you hold that definition and I can hold a definition of discrimination which says that if those test differences are not reflected in differences of job performance or educational achievement, then under the current situation we are going to have some disagreement over what might be discriminatory.

Mr. **Weiss**. Well, perhaps where we went off was in my use of the word "bias" which is what I started to use and then you gave me a double definition as to what kind of bias, how bias could be defined.

You gave me two responses and I thought that there was a distinction in your attitude as to those two, but I am not sure. Perhaps we ought to really clarify it.

Is it your view on the basis of your familiarity with the tests in the higher education sector that the same conclusions or similar conclusions can accurately be drawn, to wit, that the tests tend to, if not screen out, to demonstrate lower scores on the basis of racial, sexual, geographic, or similar kinds of differences?

Dr. **Harper**. Yes, I would—without reference to a number of specific studies in the area—I would say it is my general impression of the literature that that in fact is the case, that is, there are group differences in the educational area as well as in the occupational area. And I don't find that surprising.

Mr. WEISS. And further, do you then believe that by opening that process up as you say you find these hearings in relation to the testing process to be important and to be helpful, do you find that the similar kinds of opening up of the process for examination purposes would be helpful and beneficial in relation to tests in the postsecondary education field?

Well, let me rephrase it. I may have confused you with my verbiage. You had said just a few minutes ago that you felt that these kinds of hearings and similar kinds of hearings conducted by other committees were very helpful in trying to get a clear understanding as to why the PACE examinations, for example, seemed to be biased against certain groups or show results indicating that certain groups do not do as well as other groups.

You said that you found that that was helpful to have these hearings in relation to that particular area. My question is, then, Do you find these hearings also to be helpful in relation to the testing processes in the postsecondary education field?

Dr. HARPER. I think I would defer to the experts in that area. I think it would be presumptuous of me to speak for those who either represent the area or are experts in that area.

Our area of work and expertise is in the occupational field, the employment field, and particularly the Federal personnel and compensation.

Mr. WEISS. Are there any further questions?

Mr. Buchanan?

Mr. BUCHANAN. Thank you, Mr. Chairman.

This is as good a time as any for me to reiterate a continuing concern of mine, and that is the need for the physician to heal himself.

Legislation before us does not deal with the problem of the Federal Establishment, though the Federal Government is the Nation's largest employer. It is a white male establishment in its upper echelons, with the exception of EEOC which may approach being a black male establishment, except at leadership level at this point in history.

I wonder, in light of your testimony, if we are going to deal with the problem of what role testing may play in discrimination on an ethnic basis, should we not do something in the Federal Establishment as well, rather than deal only outside the Federal Establishment, as is the case with this legislation?

Mr. GOULD. One of the problems I am having is that the legislation deals with postsecondary education and excludes all occupational testing. In the Federal sector we are talking about basically occupational testing.

I am having difficulty drawing the analogy between the two, asking us in the Federal sector to abide by something that we are not trying to apply to the private sector.

Mr. BUCHANAN. Well, the Gibbons bill, which is also under consideration before this subcommittee does include it. Mr. Weiss' bill does not deal with occupational testing but the Gibbons bill, which is formally before us as a part of these hearings, does.

The point I am making is that we seem to make a consistent exception for the Federal Establishment. We lay requirements on the private sector, whether it is within the area of safety and health, discrimination, or women's rights, while we tend to make consistent exceptions for the Federal Establishment under some, I assume, theory that somehow the Federal Government is going to do right even when others do wrong.

It seems to me that your testimony has indicated that if there is a problem in the area of testing, that problem certainly exists in the Federal Establishment.

Mr. GOULD. I think that is a fair statement. Yes, there is a problem with testing in the Federal Establishment in the questions of whether the tests are job related and valid and that will be decided in the court under an existing case. We did find that the test does screen out large numbers of minorities.

I don't know what the problem is—it does not appear to be the test. It appears the people coming to that test have been deprived for one reason or another, of the opportunity to come to that test as equals.

Dr. HARPER. Mr. Buchanan, might I ask you a question? Would your question extend also to the military who have historically been large users of written tests?

Mr. BUCHANAN. Yes, certainly.

Well, I have, in earlier hearings, expressed that my greatest personal concern is the extent to which the test may be accurate rather than inaccurate. If these tests are in fact accurate, they reflect some much deeper problems in the system with which we also need to deal.

Thank you, Mr. Chairman.

Mr. WEISS. Thank you, Mr. Buchanan.

I wonder, just so we have something in the record about it, Mr. Gould or Dr. Harper, could you give us some background as to what in fact the congressional and/or committee history has been over the course of the last 3 to 5 years in attempting to focus on the examinations, the Federal examinations that you have been talking about?

Mr. GOULD. I think we can provide that for the record for you, sir. To my knowledge there has not been a lot. The primary focus has been as it relates to veteran's preference and the impact that the veteran's preference has, the augmented scores which prevents women from competing equally.

The work that we have done on the PACE exam did result in hearings in the House Post Office and Civil Service Committee.

Mr. WEISS. That is what I am asking you about. Can you tell us when those hearings took place, how many hearings took place, and what the results of those hearings are to date and what the projections are, if you know?

Mr. GOULD. I believe the committee hearings have recently been printed. The initial hearing was held on May 15, 1979, before the Subcommittee on Civil Service, House Committee on Post Office and Civil Service. I am not aware of any legislative proposal that has resulted from the hearings.

Mr. WEISS. But you had indicated that this report which came down on May 15, 1979, grew out of what? Were you commissioned to undertake this study?

Mr. GOULD. We initiated that review on our own because we wanted to find out if it was a problem. We had heard through various comments, public and otherwise, that this was a problem. There was not information available in terms of the makeup of the people who were taking the test.

We went through quite an extensive effort to try to identify the composition of those people. When we presented the report, as I recall, the testimony and the hearings were scheduled to coincide with the issuing date of that report. So it was some time right around May 15 that those hearings took place.

Mr. WEISS. Right, so that in fact the Post Office and Civil Service Committee has had and has under consideration, under active consideration, the subject matter of these reports?

Mr. GOULD. That is correct, sir.

Mr. WEISS. Thank you very much.

If there are no further questions, I want to thank all of you. Mr. Buchanan?

Mr. BUCHANAN. Mr. Chairman, may I ask one additional question?

Mr. WEISS. Sure.

Mr. BUCHANAN. Do you have any concerns about the disclosure provisions of H.R. 4949 creating pressures upon the local schools and teachers to teach to the test, to create pressures that might have an impact on local curriculum? As a parent that idea occurs to me.

Mr. CAMERON. NEA has a concern about teachers teaching to the test relative to standardized testing in general. We have long been on record as being concerned about that. Anytime that you have standardized testing with all the foibles involved you are going to have human beings reacting to these standardized tests in attempting to teach to them.

That is not to say that is a general principle going on around the country, but it is a concern of ours. We are very supportive of H.R. 4949 in order to open up the whole process to be sure principles are written into the standardized tests which would preclude that kind of thing.

Mr. BUCHANAN. Thank you.

Mr. WEISS. Thank you.

I want to thank all the members of the panel for your patience as well as for your testimony.

Thank you very much.

The next panel will be Dr. John T. Casteen, Mr. Frank Erwin, Dr. Edwin Fleishman, and Dr. Judy Hall.

Mr. Buchanan, I understand that you are acquainted with Dr. Hall and perhaps you would like to introduce her. I had not given any of the credits. Let me do that at this point and perhaps you will be able to expand on your credits.

Dr. Casteen is the dean of undergraduate admissions, University of Virginia.

Mr. Erwin is a member on the Advisory Panel on Declining Test Scores representing the American Society for Personnel Administration.

Dr. Fleishman is past president of the Division of Industrial and Organizational Psychology of the American Psychological Association and president of the Advanced Research Resources Organization.

Dr. Judy Hall is president-elect of the American Association of State Psychology Boards.

I see five people there. Who have I not mentioned?

Dr. Fleishman. Dr. Mary Tenopyr, who is currently president of the American Psychological Association Division of Industrial and Organizational Psychology, who is here as a resource.

Mr. Weiss. We have Dr. Casteen first on our witness list and perhaps you would like to start.

STATEMENTS OF DR. JOHN T. CASTEEN III, DEAN OF UNDERGRADUATE ADMISSIONS, UNIVERSITY OF VIRGINIA; FRANK ERWIN, MEMBER, ADVISORY PANEL ON DECLINING TEST SCORES, REPRESENTING AMERICAN SOCIETY FOR PERSONNEL ADMINISTRATION; DR. EDWIN FLEISHMAN, PAST PRESIDENT, DIVISION OF INDUSTRIAL AND ORGANIZATIONAL PSYCHOLOGY, AMERICAN PSYCHOLOGICAL ASSOCIATION, PRESIDENT, ADVANCED RESEARCH RESOURCES ORGANIZATION; DR. JUDY E. HALL, PRESIDENT-ELECT, AMERICAN ASSOCIATION OF STATE PSYCHOLOGY BOARDS; DR. MARY L. TENOPYR, PRESIDENT, AMERICAN PSYCHOLOGICAL ASSOCIATION, DIVISION OF INDUSTRIAL AND ORGANIZATIONAL PSYCHOLOGY; A PANEL

STATEMENT OF DR. JOHN T. CASTEEN III, DEAN OF UNDERGRADUATE ADMISSIONS, UNIVERSITY OF VIRGINIA

Dr. Casteen. Mr. Chairman, I have submitted a written statement to which I will refer briefly in the time allotted to me.

Attempting to primarily summarize the points covered in that statement and to indicate how I think they have to do with the practice of undergraduate admissions, especially in a State university, I have described in the statement three areas in which I and many of my colleagues in admissions have been concerned about the possible applications of H.R. 3564 and H.R. 4949 on the trade that we practice and on the students that we serve.

These areas include the predictable effects of these bills on how students progressed from high school to college and from undergraduate to graduate and professional schools, their impact on the quality of education offered to young Americans at all levels, but especially in secondary schools, and their implications for the opportunities available after high school to certain groups of young people who have only in recent years begun to participate in mainstream, postsecondary education, which is to say the opportunities available to the nonwhite, the nonaffluent, nonsuburban, the handicapped, and even now women.

The first area of my concern has to do with the possibility that these bills may disrupt orderly progress for students from high school to college. I am concerned about several possible implications in this area.

One is the potential loss of the existing system of verification that standardized testing allows us to use in evaluating other credentials that appear in the student's files.

As you know from other testimony, responsible academic admissions officers are concerned more perhaps with other parts of the record than with standardized testing, particularly with the high school transcript, recommendations from schools, statements made on their own behalf by students, and especially with their own historical experience with students from similar backgrounds and similar goals.

With that body of evidence, the standardized test scores have provided for us an element of predictive validity that has made the system become extremely effective in identifying students who can indeed succeed, especially in highly demanding, academically competitive programs.

We are concerned that the legislation appears to us to disrupt this system of verification without providing a reliably documented replacement system on which we can rely to provide the necessary information to make responsible selections.

We are concerned also that if this disruption occurs, the least harm will occur to those students who come from the most privileged sections of our society and the most harm will come to those students who come from the least privileged sections, especially from those who are, as I said before, nonwhite, nonaffluent, nonsuburban, et cetera.

We are concerned, second, about the possibility that the enactment of such legislation as that contained in these two bills may actually serve to interrupt an ongoing drive toward reassessment of the quality of schooling that goes on in this country.

In my written remarks I have referred to three signal documents in this process of reassessment.

The first of these documents was the announcement in 1975 by the College Entrance Examination Board that SAT scores since the year 1963 had declined in all the relevant areas of measurement for all significant parts of the population, cutting across geographical, socioeconomic, ethnic and other lines.

The second is a series of articles published in the summer of 1976 by the Los Angeles Times about the declining American education, articles which attempted to quantify the causes for the declining scores and to begin a national climate in which we could begin a discussion about the possibility of schooling to our students.

The third document is the report of the Wirtz Commission, the panel trying to diagnose the score decline in terms of the experience of students prior to taking the tests.

Since these documents became available to the public, the developments have seemed to me to be a mixture of good and bad. For example, I have seen no major assumption of leadership in the direction of reassessing schooling on the part of the major professional associations, including the NEA, AFT and other organizations representing teachers.

We have seen little from Government in the way of addressing the actual causes for declining measures of achievement at the end of high school and at the end of college. We have seen movement on the part of local jurisdictions, on the part of individual schools.

We have seen a growing sense that the public is alarmed that students seem not to be able to read as well as they should be able to, to deal with the English language, write or compute as well as they should be able to, and moreover in all fields to perform as well as they could before the educational revolutions of the sixties and early seventies.

We see evidence that school boards, private employers, the Government, the military, find all of our students suffering in one respect or another, but many of our students, especially those who are nonwwhite, nonaffluent, et cetera, receiving at the end of high school credentials that appear not to present the qualifications necessary to make a meaningful contribution to society after school.

Testing in effect blew the whistle on the failure of the school system. We are concerned that legislation aimed at interrupting the normal flow of information from properly evaluated testing does not address the crime but punishes the police officer who blew the whistle.

Third, we are concerned about the possible interruption of advancing educational opportunities to constituencies which have historically represented the mainstream of education in this country.

Testing in the last 5 years has diagnosed areas in which schooling and society in general have failed to provide the opportunities they should. It is not the fault of testing. If we do poorly in teaching black children to read, to write, to think in the English language and to compute, it is instead the function of testing as it has to point out areas of failure and it is after that particular demonstration the function of Government and of education to move expeditiously and intelligently in resolving the problems that have turned up.

Testing has helped us to bring new constituencies into the educational mainstream. I need only point to the progress of women in receiving entry into the fields of engineering, business, architecture and other fields that have historically been closed to women and, second, not all black students score lower than white students and not all students who have been outside the mainstream are suffering because of test scores, and that in fact the obvious deficiencies in the school programs pursued by such children, a fact I see daily in my work when I find bright black children who have not been advised to pursue studies in the social studies, engineering, mathematics, science and other areas, in seeing that evidence in my own daily work, I am alarmed by the prospect that the normal flow of information that tells us these children may have not been well served may be interrupted by legislation of this kind.

Testing has in fact in summary served to break certain stereotypes. It has told us that women can function well in advanced programs in the sciences and in mathematics, that many black children, especially children who have received responsible and careful attention in the lower and secondary schools, can in fact succeed even in the most collective and competitive college situations and can go on to become professional leaders of real substance.

The largest concern is this: For the present system to be properly indicted, to be shown not to serve this country well, it seems to me the case against it must be made on a level of integrity and with the kind of comprehensiveness that I have not seen in the materials previously submitted to this committee.

It seems to me, second, that if the present system is to be replaced, the replacement system must be documented as being equally valid and must be directed to the concerns of the nonmainstream student.

My caution to the committee is that we as a country must address problems that have now been diagnosed with sufficient clarity that we know where we must begin to move.

It seems to me the largest mandate is not toward controlling testing but instead toward reexamining the quality of educational experience and seeing to it that the very young people about whom we are most concerned in bringing them into the mainstream do in fact receive the kind of background necessary to perform well, not only on the tests but also over the course of their academic and working careers to achieve the kind of leadership and success that makes for responsible citizenship.

[Prepared statement of John Casteen follows:]

PREPARED STATEMENT OF JOHN T. CASTEEN III, DEAN OF ADMISSIONS, UNIVERSITY OF VIRGINIA

I am John Casteen, Associate Professor of English and Dean of Admissions in the University of Virginia. I have come today to express concern about what I believe would be the adverse effects of passage of H.R. 3564 and H.R. 4949 (as now written) in three areas: their predictable effects on how students progress from high school to college and from undergraduate to graduate and professional schools; their impact on the quality of educational experience offered to young Americans at all levels of education, but especially in secondary schools; and their implications for the opportunities available after high school to certain groups of young people who have only in recent years begun to participate in mainstream post-secondary education, which is to say to the non-white, the non-affluent, the non-suburban, the handicapped, and (even now) women. I speak from my experience as a teacher and admissions officer, especially from experience since 1975 as Dean responsible for undergraduate admissions at a state university that attracts able applicants from throughout the country and from virtually all social, economic, and ethnic backgrounds. My own background is in public education. I share concerns expressed to you previously by Fred Hargadon of Stanford University, but I want today to speak primarily to the implications of these two bills for public higher education in times of declining student populations, declining net indicators of academic accomplishment, declining real and relative resources to support education, and mounting public concern about the final results of education.

The potential obstacles posed by H.R. 3564 and H.R. 4949 to orderly student progress from one level of education to the next have, I believe, come previously to the Committee's attention in testimony offered in behalf of the American Council on Education and the Educational Testing Service. With other admissions officers with whom I have discussed these bills and their predecessor in the state of New York, I share the A.C.E.'s belief that public disclosure of currently secured test items, however well-intended, must inevitably diminish the quality of information available to academically selective colleges when they review applications and also the credibility of the credentials available to students who seek to establish the viability of their preparation for further study. Having seen no practical alternative to the existing system of secured, standardized tests designed to measure the final products of education at each level, many of us in admissions fear that H.R. 3564 and H.R. 4949 may, especially in the first few years following their enactment, hinder us in our efforts to understand the complete package of credentials presented by applicants, and therefore force us to make less reliable decisions than we can now make while simultaneously depriving students of the kind of reliable validation of high school courses and grades that now comes to us in the form of standardized test scores based on secured and validated tests. We are concerned that the reporting and disclosure requirements of these bills, combined with the necessity that the test preparation agencies develop massive new batteries of test items and heretofore untested methods to equate scores and validate their meanings, will drive costs beyond what most students can pay. We suspect that students from privileged communities who apply to the most prestigious colleges will suffer least from rising costs and sinking test credibility, but that less affluent students will be increasingly less able to pay the cost, hence less likely to qualify themselves for consideration by the major colleges. Well-known secondary schools, schools in our more affluent neighborhoods, which are typically white and suburban, and schools with established traditions of preparing students for the most competitive colleges will continue to succeed as they always have. By contrast, inner city schools, rural schools, schools whose students happen not to be like those who commonly qualify for admission to the most competitive colleges, will almost certainly suffer if the final measures now available become suddenly less available and less credible.

I share the conviction of the American Council on Education that standardization is a necessary condition of existence in contemporary educational testing. This conviction grows not out of personal preference (who can favor gridded answer sheets, machine scoring, or the intense atmosphere of the rooms in which tests must be administered?) but out of the practical requirement that students be able to demonstrate that their preparation is genuinely comparable to that offered by others, including students of varying backgrounds, geographical origins, and school experiences. The size of our population, the diversity of our educational system, and the rich variety of backgrounds presented by applicants together demand that we maintain a workable system to verify outcomes. Not all schools teach the same things; not all students learn the same things; not all colleges need the same kinds of preparation in their freshmen. But one must have reliable means to sort out what credentials prove. The Ivy League schools, major national institutions like Stanford or the University of Chicago, and a handful of highly selective public institutions, perhaps including the University of Virginia, can if we must develop unique, in-house, non-standardized entry examinations to permit us to validate secondary school credentials and to predict student success in our own curricula. No doubt we will do so if genuinely reliable national testing disappears from the scene. We had such tests before the College Board was formed. But the cost, both in dollars and in lost opportunities for students to participate in the most rigorous kinds of education available in our colleges, will be high. As student populations shrink, as more colleges undertake policies of open enrollment, hence taking less interest in the adequacy of high school preparation and perhaps also in the integrity of their own programs, indeed as colleges realize that the demand for their product, education, is less than it once was, concern for quality will seem less important than will filling the dorms, paying the faculty, and staying in business. Analogously, students seeking to progress from less prestigious undergraduate programs to the major national graduate and professional schools will suffer, as New Yorkers may next year, if the graduate admissions tests become less reliable and more costly than they are now. Students at all levels need properly devised, generally administered, readily comparable final credentials of the kind now available through the tests.

My concern about the potentially adverse effects of H.R. 3564 and H.R. 4949 on educational quality, especially in high schools but also in colleges, grows out of experience since 1975 with evidence to the effect that schools have been in recent years less successful in teaching reading, verbal reasoning, writing, and arithmetic than they ought to be and once were. The evidence comes from sundry sources. It first became available to the public in 1975 when the College Board revealed that mean S.A.T. scores had declined precipitously in the years since 1963. The College Board's announcement came as no surprise to many of us in teaching and admissions, because we had known for several years that something was amiss with our students. At the University of California in Berkeley, where I was teaching when the College Board announcement appeared, we had by 1975 been engaged for several years in major remedial efforts in the area of writing. More than one classroom teacher had already modified his or her methods and expectations in order to deal with new sets of student needs and deficiencies, especially in the most basic disciplines. National discussion of the score decline began at once. Many commentators condemned the College Board for making the announcement on grounds that it tended to bring discredit on schools; others commended it on grounds that it proved that times were bad; still others attacked testing itself, alleging that the tests, not student performance, were growing worse. This discussion gained focus with the publication in August 1976 of three feature articles in the Los Angeles Times. Written by Times education writers Jack McCurdy and Don Speich and entitled "The Decline of American Education," these articles presented for the first time empirically sound evidence that the apparent decline in student skills was "real and not, for example, the result of changes in the ethnic or socioeconomic makeup of test takers." The Times permitted other newspapers to reprint the series, and distributed copies to educators, members of Congress, and other concerned persons.

McCurdy and Speich offered evidence that the decline was by no means restricted to SAT and ACT scores, that indeed it appeared with disturbing consistency in virtually all available indicators of educational progress and at all levels of education. They also suggested that declines in final measures of achievement closely matched the upward progress of the mean grades awarded to students during a decade of grade inflation and lowered requirements. Seeking causes for the decline, they suggested among other factors changing school curricula, especially in high schools, changing student populations, changing educational objectives, new ways of learning (notably television learning), and progressive deterioration in the fabric of the American family. McCurdy and Speich dealt harshly with our schools; their third article made the case that education generally, and especially secondary education, was doing little or nothing to remedy the ills of our young people. Together, these articles became and remain major documents in the case against our schools as they now function.

These articles were followed in turn by the creation of the Wirtz Commission, properly known as the Advisory Panel on the Scholastic Aptitude Test Score Decline, sponsored jointly by the College Board and the Educational Testing Service. *On Further Examination*, the panel's summary report, along with its companion appendices, ought in my judgment to be required reading for all framers of educational policy, concerned citizen leaders, and above all teachers. Without attacking schools or school leaders, *On Further Examination* documents at least seven major contributors to the score decline: the changing composition of the college-bound population (a cause generally acknowledged to be consistent with our national purposes and of no immediate concern in a negative sense), the proliferation of elective courses in schools, especially in English and verbal skills, where student performance has declined most precipitously, a diminished seriousness of purpose in the learning process as reflected in tolerance of excessive absenteeism, grade inflation, easier textbooks, and generally lower standards, increased viewing of television, changing structure of the family, social turbulence of the sixties and seventies, and diminution of learning motivation.

The College Board became involved in this discussion of the causes for the score decline only after serious and thorough consideration of the possible outcomes of its involvement. A desirable outcome might have been serious national scrutiny of the purposes and methods of American education at the three-quarter-century mark, with an eye toward rediscovering what worked well in the past and either adapting it to present circumsntances or developing effective new solutions to our problems. Sadly, neither Congress, the Office of Education, state leaders, nor the major educational associations undertook this reassessment. A second possible outcome was a reaction against the tests. Such a reaction is, I fear, best exemplified by legislation that would vitiate the system of final measures, the existing standardized tests, without guaranteeing the immediate development of equally effective, equally inexpensive, and equally general final measures. I take it as proper and constructive that our political leaders should promote the use of good measures of educational accomplishment, and also that Congress should monitor the services delivered nationally to students, colleges, and American society by the test preparation agencies. Yet one does not cure disease by depriving the diagnostician of his tools or reverse bad news by executing the messenger who brings it. The educational establishment failed to take effective action when students first showed signs that the education revolution of the sixties had failed. Instead, the testing agencies and the College Board, the keepers of the final measures, had to do so. Without such agencies and without their conviction that the public must know when the evidence suggests that students are not learning what they should, we might not know even now just how desperately we need to undertake a thorough reexamination of how we raise our children.

My third concern, extending educational opportunity to broader constituencies, especially to persons not historically within the mainstream of higher education, has been discussed previously by spokesmen on both sides of the issues raised in H.R. 3564 and H.R. 4949. Two matters critical to the actual practice of admissions seem to me not to have received proper analysis. One is the place of testing in the admission process. In colleges that deal responsibly with their applicants, the SAT and similar tests serve two functions: they allow us to understand the credentials of individual applicants in light of norms that are larger than those of the individual high school from which the applicant may come or of our own student population, norms that are both diachronic (because the test scores equate across time) and synchronic (because all applicants in a given year have faced essentially the same measure of their readiness to perform in our freshman courses): and they allow us to make reasonably close assessments of student skills in the basic areas—in reading, vocabulary, writing, and arithmetic. For every college whose selection system I know, other credentials matter more; indeed, the test preparation agencies themselves enjoin colleges against simplistic reliance on test scores, and call on them to develop what have come to be complex models of student preparation as it can be assessed in sundry ways. High school transcripts tell us how rigorous a program of studies an applicant has pursued and how well the student compares to other applicants from his or her school and from schools proved historically to be similar. Counselor or teacher recommendations, less readily available now than before passage of the Buckley Amendment, can add insights into student motivation, ability to cope with obstacles, commitment to good citizenship, and other relevant matters. Student self-descriptions can detail personal goals, non-school accomplishments, and a wealth of other topics, while also giving us a direct sample of how effectively applicants write and think. Each element matters if the admissions office does its job conscientiously and well; obviously the non-quantitative measures matter most to those of us who deal with large applicant pools and seek to admit students whose future accomplishments, academic and other, are likely to be well above average. The peculiar value of the standardized test score in the selection process is its historical and contemporary comparability, its potential to indicate whether a student from a specified background has learned what he or she must in order to compete and succeed in the requested program of study.

The second matter is the application of test scores in the actual process of selecting students whose backgrounds differ from those of what one might call this generation's mainstream college student, especially in the more selective colleges. Women, non-white students, handicapped students, older students, students from rural backgrounds or from the kinds of inner city schools that have developed since World War II—these students ought properly to be the subject of continual study on the part of all admissions offices. Each of these groups has emerged in relatively recent years as a new student constituency in colleges that previously enrolled primarily white male students from our more affluent neighborhoods. In certain instances, the S.A.T. and similar tests have helped to break down barriers by proving that suspected or alleged deficiencies simply do not exist. Women who have entered programs in engineering, business, and architecture have benefited from the availability of evidence that bias against them lacks an empirical basis. In other instances, test scores have proved to be less reliable because normative student performance either fails to predict college grades accurately or falls outside the limits within which we can equate scores from our history files. In the first year or two of my effort to increase the number of black students enrolled at the University of Virginia, I came to the realization that no predictor, including test scores, told me what I needed to know in order to predict success. Like admissions officers at virtually all other highly selective colleges and universities, I developed new ways to analyze the credentials of our new student populations, and I developed history files based on this new experience.

Anyone who works with test results knows the phenomenon labeled cultural bias. Actually, cultural bias is not one phenomenon, but several, and it is by no means restricted to standardized tests. It appears in studies that show that black students perform in manners unlike white students (although the statement itself is too general to be reliable), that men and women have different areas of strength, that urban, suburban, and rural youngsters mature at different rates, or that students from different parts of the country know and know how to do different things at the end of high school. It includes also the elementary fact that tests, like books, newspapers, and even the presentations made to this committee, must conform to some linguistic norm. Spokespersons for the test preparation agencies have, I believe, testified previously about their efforts to minimize the potentially deleterious effects of cultural bias; schools, employers, and government agencies struggle with related problems every day. With regard to the use of testing in the admissions process, certain principles must hold true: scores are useful only when their value as part of the complete package of credentials has been reliably validated; students, especially those seeking entry from non-mainstream backgrounds, deserve the benefit of the doubt when the institutional history file will not support a responsible judgment. To put the matter into practical terms: no responsible admissions committee can deny a black applicant merely because his or her test scores differ from those commonly found in the mainstream population; nor can a responsible selection agent in any other area of competitive endeavor justly refuse an applicant because he or she does not look like other applicants. College admissions officers who do their homework have known these truths for years; indeed, the College Board and other such agencies have throughout my own career worked to educate the profession on how best to assess the promise of non-mainstream applicants. Ironically, the agencies that I know to use tests irresponsibly, which is to say mechanically, without regard to other indicators that can be documented to have more validity in predicting student performance, are not college admission committees (although no doubt more than one errs) but governmental agencies, state and federal, that ignore the College Board's advice by using fixed cut-off scores to determine eligibility for scholarships or entry into special training programs. H.R. 3564 would require that test takers be notified of "the score required to pass the test . . . or the score which is generally required for admission to institutions of higher education." No one passes or fails tests like the S.A.T. and A.C.T.; nor is any score generally required for admission to institutions of higher education. H.R. 3564 amounts in this respect to a mandate for irresponsible selections, especially selections harmful to the interests of non-mainstream students.

My largest reactions to the proposals embodied in H.R. 3564 and H.R. 4949 are these: many actions mandated by these bills are already common practice among the major test preparation agencies; others are not common practice but perhaps should be—the presentation in behalf of the American Council on Education addressed these matters in detail; still others derive from faulty premises about how tests fit into the admission and selection process, about how students mature, and especially about the relation of testing to educational and social policy. Tests developed in this country because colleges and students needed the final measures implicit in them. They developed as results of private endeavor, and heretofore they have performed their service to students and to college without expense to government and without government's having developed expertise of the kind requisite if government is now to define reasonable standards of performance for test makers and users. With most of my colleagues, I welcome your examination of the role of testing in the selection process. I believe that education generally expects to comply in every sense with any legislation that may develop from your deliberations. At the same time, I am concerned for our students to find in these bills so many potential hazards and so little concern for the beneficial results of the system that has developed. My judgment is that education needs reformation, and that citizens, educators, and students want to see a deliberate national move toward more effective schooling for all young people, especially for the non-white, the non-affluent, the non-privileged. To that end, legislation that risks sacrificing the effectiveness of the existing tests as measures of where we are and of where we need to go seems to me to fall far short of the mark. Congress should and can provide leadership in the search for an end to the progressive declines in student performance in the basic skills, for an end to the ancient disparity between the educational outcomes of our most privileged citizens and those of citizens who have yet to participate fully either in higher education or in the process of self-government. Sad to say, H.R. 3564 and H.R. 4949 seem to me to ignore our most compelling needs while taking away our ability to develop and employ at least one means of making education effective, meritocratic, and in the end democratic.

Mr. **Weiss**. Thank you very much.

We will commence questioning after all the members of the panel have spoken.

Mr. Erwin?

STATEMENT OF FRANK ERWIN, MEMBER, ADVISORY PANEL ON DECLINING TEST SCORES, REPRESENTING AMERICAN SOCIETY FOR PERSONNEL ADMINISTRATION

Mr. **Erwin**. Mr. Chairman and members of the committee, my name is Frank W. Erwin. For the past 11 years I have been president of Richardson, Bellows, Henry, & Co. a nationally recognized personnel research firm headquartered here in Washington, D.C.

In that position I have been involved intimately and directly in the development and validation of selection procedures, particularly tests, and during two of those years served as member of the advisory panel on the scholastic aptitude test score decline.

I am appearing before you this morning in the hope that what experience I have will be helpful to you, and additionally to express the views and the concerns of the American Society for Personnel Administration (ASPA) on the proposed testing legislation you are considering.

This society, which consists of over 23,000 personnel practitioners in business, government, and education has cause for concern since any action of this committee which in any way compromises the quality of those entering higher education ultimately will compromise the quality of those who we expect to manage our enterprises in the years to come.

In that context, we have approached this testimony with a great sense of apprehension which we hope this committee will share. H.R. 3564 and H.R. 4949 venture into one of the most complex fields ever addressed by this committee and you are being asked to take a precedent setting and dangerous step toward regulation of the educational admissions process and the procedures which serve, at least in part, as indicators of the quality of our society and its institutions.

The statement which we make today, therefore, attempts to address major elements which must be considered as part of your determinations as to the need, the purpose, and the potential impact of the proposals before you.

I think my predecessor to my left has stated it quite well in terms of the role of tests.

In Excellence, John Gardner wrote that, "Tests are designed to do an unpopular job." An untutored observer listening to critics lash out at the imperfections of the tests might suppose that the criticisms would be stilled if the tests were perfect. Not at all. As the tests improve and become less vulnerable to present criticism, the hostility to them may increase.

A proverbial phrase indicating complete rejection is, "I wouldn't like it, even if it were good." With the tests, the more appropriate phrase might be, "I wouldn't like them, especially if they were good."

Tests are designed to serve a performance measure related to some future point in time. That is what Mr. Gardner refers to. The concept he expresses, however, also is applicable to the role of test scores as messengers about our excellence as a society and the extent to which our institutions are functioning.

Interestingly, while you are being asked to begin restricting the use of tests, the messages we have been getting are increasingly unflattering and unpleasant.

In education, these messages center around the fact that our schools, particularly those where our poorest citizens are clustered, are doing an increasingly inadequate job, despite massive transfusions of Federal funds. Fewer and fewer students are able to read and write, or add, subtract, divide, and multiply, and the National Education Association's response is a call for a moratorium on the use of tests.

In the employment setting, due principally to civil rights enforcement activities, tests and test practices are better than ever before and these messages are centering around the fact that those who do less well on the tests, irrespective of subgroup, whether that be black, white, rural, urban, female or male, also are doing less well on the job.

All of this should raise questions, first concerning how we can measure quality objectively without the use of tests, and second, whether some of the supporters of H.R. 3564 and 4949 have a hidden agenda.

The test messengers, like thermometers, are telling us that we have a worsening illness. Rather than seeking a cure for the disease, those for whom the test results are most threatening would ask you to restrict, and preferably destroy, the use of thermometers. Like Persian kings, they would slay the messengers who bring them bad tidings.

You were asking earlier about bias in tests. That is an inaccurate or incomplete measurement. The measurement is how varying groups score on it, but most importantly, does that set of scores predict subsequent performance.

Bias in tests and bias in items, different groups performing differently, is not real measure of whether a test is indeed discriminatory or appropriate for use. The measure is: Does the difference in test performance correlate with differences in academic or occupational performance?

It is not easy to fix what we are finding with differences in test scores. There have been numerous experiments to try. We found, for example, removing test items which appear to differentiate between groups, in general ability tests, to take that procedure would more than likely result in the removal of virtually all of the items in the tests because the difference is consistent across all of the items.

Those who have supported these bills have diligently resisted or downgraded all references to the tests' validity and reliability. Rhetoric has been used in place of evidence and misleading concepts such as extent of variance accounted for have been waved like flags at a parade.

As a firm rule, given the critical nature of these proceedings, this committee should remain unconvinced by such arguments and should know that a wealth of data establishing the validity and usefulness of these instruments for their purpose is available to the committee, as it was to the advisory panel and has been to responsible researchers. Contrary to what you have heard, data has been available.

There is, too, a substantial number of studies available concerning the extent to which these tests also are applicable to subgroups within the total test taking population and whether changing the content, the speed, or the format of the tests eliminates the deficiencies which exist among certain subgroups, generally, no, it does not.

In general terms, the committee first should be wary of any statements based on evidence from one study. Responsible professionals do not use one study to persuade, particularly when there is a larger body of evidence available, as there is here. To do so is contrary to professional ethics and is analogous to taking a remark out of context.

Second, the larger body of evidence shows that there is a consistent and likely underestimated relationship between test scores and subsequent performance which is sufficiently high to justify the use of the tests. Obviously tests are far from perfect predictors, a condition which is even more true of any subjective approaches which you have heard mentioned here.

If you were to summarize the data that exists, it would say the tests are valid for their purposes. High school grades are generally better predictors, but the use of tests in combination with the grades are a better predictor of college performance.

High school grades alone, you have heard questions about if high school grades are as good as or better than tests, why not use grades. Those who raise the question typically share a misplaced conviction that tests are not valid for minorities or females. The fact is they are.

The proponents of this argument are not aware of the validity of test scores and high school grades used in combination. By recommending using just high school grades, you are recommending a less valid means for selecting people for these critical functions and aspects of our society.

It may be that high school grades would not show such a relationship if test scores were not present to interact with them. It is highly likely that colleges and universities are using test scores to a significant degree to determine the reliability of grades, for it is senseless to argue that A's and B's from District of Columbia schools are equivalent to those of Montgomery County.

I think a poll of the membership of both Houses in terms of Representatives who live in the District of Columbia and who have children of school age, I would ask, where do you send your children. It is not likely to be the District of Columbia public schools. They have made their judgment the same as institutions make their judgments.

If the proposed bills are passed and test usage diminishes, as it will, and don't fool yourselves about that, college and university admissions systems will rely more heavily on high school grades, but more in terms of "safe" schools—those high schools known to be of higher quality.

The net result could be fewer students selected from poorer schools, particularly those in the cities with their higher proportion of minorities. With the objectives and pressures that the admissions systems have, candidates from the Montgomery schools would be the safer bets and the safe choice would more often than not be made.

You have asked about test scores and income level as if the relationship between test scores surprised you. It should not. The relationship exists between test scores and income levels. But that is a simplistic look at it.

A fuller exploration of the subject would show that families with higher income include more two-parent families, more higher educated families, smaller family size, more value being placed on education, more books in the home during the early formative ages, and better diet for both mother and child in the early years, along with attendance typically at better secondary education institutions.

All of these things are related to higher scores and higher performance in school. It also should be clearly understood, especially to those who have raised the question as to whether high school grades would suffice as a sole measure of academic success, that high school grades also show these kinds of relationships.

H.R. 4949 requires testing agencies to provide statements concerning the extent to which test preparation courses improve test subjects' scores. This is an impossible and meaningless task. It would be easier to ask beauty schools to report on the extent to which they make change. With test schools it is virtually impossible to do.

And the question must be asked, does it have any meaning, what meaning is increased if it occurs? There can be no question in my mind despite some claims to the contrary that have been made before you that coaching makes change, I think. There is no question, too, that coaching is a burgeoning industry.

I think, Mr. Weiss, you have pointed out the number of branches of just one organization in the test coaching business. Passage of the bill, however, will add to that growth and make the coach's job more easy. They don't have to resort to circuitous means to get that information.

There is still considerable controversy on the extent to which coaching changes scores, and more importantly, on the real value of the changes, if indeed they do occur.

It is reasonable to assume, however, that coaching for most persons will result in an increase in the score they otherwise would have gotten. It is difficult to measure the magnitude of the difference and the extent to which other variables will affect the differences which will occur.

But it is simplistic to evaluate coaching in those terms. Based on studies conducted on other elements, including the performance of various subgroups on tests containing content and more related to their subgroup and the effect of the removal of time limits on speeded tests, it is reasonable to expect that majority enrollees who are already higher scoring will in such courses have a score gain at least as much, perhaps more, than minority enrollees.

If one considers all of the things related to higher income that were discussed above, it also is reasonable to expect that enrollees in these courses from higher income families will achieve changes at least as high and probably greater than low-income families.

In short, it is unlikely that establishing funding to insure that all persons have access to coaching schools will do much more than wildly enrich the coachers. Those who would have scored higher without the coaching will be more likely to take advantage of it and will achieve greater gains. The differences, therefore, will not be erased and may well be increased.

If there is to be a test of test coaching, reporting score differences that occur is not sufficient. The test of test coaching will be its validation, to what extent do those persons entering test coaching schools, and who have a subsequent increase in scores, perform to a similar higher degree when they go into the educational setting or is that increase false?

For example, coaching on analogies will improve analogy-solving skill and those coached will do better on that item type.

The question remains as to whether that improved test-taking skill improves performance in college. The solving of analogy problems for some reason is related at least in part to college grade achievement. It is, however, not a direct sample of the performance behavior; that is, getting good college grades.

Since the relationship is not clearly direct, a coached ability to solve analogy problems may reduce or eliminate the predictive power of that kind of item for the individual involved.

In short, the difference may be artificial in terms of test ability but have no meaning in terms of test performance in the academc setting. Test coaching and the true test of it lies in the validation of increased performance or the performance as a result of coaching to determine empirically, does coaching and the increases resulting from it result in a corresponding increase in the performance of those individuals coached in the college setting?

Mr. WEISS. We have about 7 or 8 minutes to answer the quorum call. We will have to leave, but we will be back as quickly as we can.

[A brief recess was taken.]

Mr. WEISS. We are ready to resume.

I should indicate and should have at the outset, at any time if any of you wants to summarize your statement or highlight it, the entire statement of each of you, without objection, will be entered into the record so use your own judgment if you want to have it entered into the record automatically, and how much you want to give orally.

Mr. ERWIN. I will close very quickly, Mr. Chairman, with several more points.

One is destroying tests, whether it will or won't. The legislation will indeed destroy some tests, particularly those in the achievement area and especially those such as autobiographical questionnaires, which I might note for you that if there is any promise on suitable alternatives, autobiographical information is probably the best promise or has the best promise.

My statement describes this system much more fully, but providing answers to questions on autobiographical questionnaires will destroy them because of the inherent characteristics of that kind of instrument.

There has been considerable research that indicates that these do indeed have less impact than the kinds of instruments we are discussing here. Achievement tests, depending on how specific the knowledge measured, will become less valid in the short term, and in the long term could eliminate quite a few of them.

The time involved to handle what is before you is extraordinarily short. The panel, of which I was a member, the advisory panel on score decline, took 2 years to reach its conclusion and even at the end of the 2 years we felt very strongly we could have used another 2. We commissioned 37 separate studies as part of that work.

The Federal agencies in publishing the uniform guidelines on employee selection procedures; namely, the Departments of Labor and Justice, the Department of Treasury involved peripherally. The EEOC and the Office of Personnel Management took 6 years to arrive at its uniform guidelines. You are being asked to proceed to mark up in some 6 to 8 weeks the subject being so complex and so critical which cannot be handled or dismissed in so short a period of time.

The opposition that you are hearing is more than from test publishers. The American Psychological Association registered its dissent in the proceedings on the New York bill. The colleges and universities uniformly oppose the bills before you.

The business community opposes the bills before you; and insofar as ASPA is concerned, it would oppose the bills in their present form and would recommend either rejection or, at the minimum, a putting on the shelf for a month. You could raise a question as to which is better for the industry, and for what we are trying to do in measurement, the threat of the iron or the iron itself.

In terms of the irons that are proposed, they are not adequate for their purpose. We do have legislation in New York. We have less onerous legislation in California.

Two States have turned down similar proposals. The National Academy of Science has been taking hearings on the subject of ability testing and is scheduled to release a report in the middle 1980's. I would recommend at a minimum that we wait until experience is gained in New York and California and the balance of the States, and at least until the National Academy of Sciences has published its report.

I think in the interim that this committee has done an extraordinary job of surfacing what is indeed a problem and that surfacing is going to have some effect in terms of things being done better and things changing.

Perhaps we could convince the Federal Government to do more than it has. Its action thus far has been to substitute for the Federal service entrance examination a similar example called the PACE.

It has halitosis in one, bad breath in the other. They are interchangeable, and there needs to be much more done in terms of suitable alternatives. The action of this committee thus far, the threat of the iron will help considerably but in terms of final action the committee should wait.

Thank you very much.

[Prepared statement of Frank Erwin follows:]

PREPARED STATEMENT OF FRANK W. ERWIN ON BEHALF OF THE AMERICAN SOCIETY FOR PERSONNEL ADMINISTRATION

Mr. Chairman, Members of the Committee, my name is Frank W. Erwin. For the past eleven years, I have been President of Richardson, Bellows, Henry, and Company, a nationally recognized personnel research firm headquartered here in Washington, D.C.. In that position, I have been involved intimately and directly in the development and validation of selection procedures, particulary tests, and during two of those years served as member of the Advisory Panel on the Scholastic Aptitude Test Score Decline. I am appearing before you this morning in the hope that what experience I have will be helpful to you, and additionally to express the views and the concerns of the American Society for Personnel Administration (ASPA) on the proposed testing legislation you are considering. This Society, which consists of over twenty-three thousand personnel practicitioners in business, government, and education, has cause for concern, since any action of this Committee which in any way compromises the quality of those entering higher education ultimately will compromise the quality of those who we expect to manage our enterprises in the years to come.

In that context, we have approached this testimony with a great sense of apprehension, which we hope this Committee will share. H.R. 3564 and H.R. 4949 venture into one of the most complex fields ever addressed by this Committee and you are being asked to take a precedent-setting and dangerous step toward regulation of the educational admissions process and the procedures which serve , at least in part, as indicators of the quality of our society and its institutions. The statement which we make today, therefore, attempts to address major elements which must be considered as part of your determinations as to the need, the purpose and the potential impact of the proposals before you.

The Role of Tests

In Excellence, John Gardner wrote that,"Tests are designed to do an unpopular job. An untutored observer listening to critics lash out at the imperfections of the tests might suppose that the criticisms would be stilled if the tests were perfect. Not at all. As the tests improve and become less vulnerable to present criticism, the hostility to them may increase. A proverbial phrase indicating complete rejection is, 'I wouldn't like it, even if it were good.' With the tests, the more appropriate phrase might be, 'I wouldn't like them, especially if they were good.'"

Tests of course serve as predictors and it is this function to which Mr. Gardner principally refers. The concept he expresses, however, also is applicable to the role of test scores as messengers about our excellence as a society and the extent to which our institutions are functioning. Interestingly, while you are being asked to begin restricting the use of tests, the messages we have been getting are increasingly unflattering and unpleasant. In education, these messages center around the fact that our schools, particularly those where our poorest citizens are clustered, are doing an increasingly inadequate job, despite massive tranfusions of federal funds. Fewer and fewer students are able to read and write, or add, subtract, divide, and multiply and the National Education Association's response is a call for a moratorium on the use of tests. In the employment setting, due principally to civil rights enforcement activities, tests and test practices are better than ever before and these messages are centering around the fact that those who do less well on the tests, irrespective of subgroup, also are doing less well on the job.

All of this should raise questions, first concerning how we can measure quality objectively without the use of tests, and second whether some of the supporters of H.R. 3564 and 4949 have a hidden agenda. The test messengers, like thermometers, are telling us that we have a worsening illness. Rather than seeking a cure for the disease, those for whom the test results are most threatening would ask you to restrict, and preferably destroy the use of thermometers. Like Persian kings, they would slay the messengers who bring them bad tidings.

The Validity of and Reliability of Tests

Some of the most cavalier testimony that this Committee has been subjected to concerns the extent to which admissions tests achieve their basic purpose; that is, to provide one objective means by which an individual's probable college performance, particularly first year grades, can be estimated. Questions also have been raised as to whether the tests show validity for all groups who must take them.

Those who have supported these bills have diligently resisted or downgraded all references to the tests' validity and reliability. Rhetoric has been used in place of evidence and misleading concepts, such as extent of variance accounted for, have been waved like flags at a parade. As a firm rule, given the critical nature of these proceedings, this Committee should remain unconvinced by such arguments and should know that a wealth of data establishing the validity and usefulness of these instruments for their purpose is available to the Committee, as it was to the Advisory Panel and has been to responsible researchers. There is, too, a substantial number of studies available concerning the extent to which these tests also are applicable to subgroups within the total test-taking population and whether changing the content, the speed, or the format of the tests eliminates the deficiencies which exist among certain subgroups.

In general terms, the Committee first should be wary of any statements based on evidence from one study. Responsible professionals do not use one study to persuade, particularly when there is a larger body of evidence available, as there is here. To do so is contrary to professional ethics and is analogous to taking a remark out of context.

Second, the larger body of evidence shows that there is a consistent and likely underestimated relationship between test scores and subsequent performance which is sufficiently high to justify the use of the tests. Obviously tests are far from perfect predictors, a condition which is even more true of any subjective approaches which you have heard mentioned here. Our own summary of all of the data reviewed would be that admissions tests are valid for their purpose, that high school grades are generally better predictors than the tests, and that use of a combination of tests and high school grades results in a more valid and therefore fairer measure of college potential, irrespective of subgroup, than either element separately. The question is, does this Committee want to propose legislation to require that form of use?

High School Grades Alone

One question you have heard raised is that if high school grades are as good as or better than tests, why not just use grades?

First, those who raise the question typically share a misplaced conviction that tests are not valid for minorities or females. The fact is they are. The proponents of this argument also are deliberately ignoring or are unaware of the greater validity of test scores and high school grades used in combination. They are in effect recommending a less valid and therefore less effective means to identify those most likely to keep this country competitive. It may well be, too, that high school grades would not show such a high relationship if test scores were not present to interact with them. It is highly likely that colleges and universities are using test scores to a significant degree to determine the reliability of grades, for it is senseless to argue that A's and B's from District of Columbia schools are equivalent to those of Montgomery County.

Finally, and most importantly, if the proposed bills are passed and test usage diminishes, as it will, college and university admissions systems will rely more heavily on high school grades, but more in terms of "safe" schools--those high schools known to be of higher quality. The net result could be fewer students selected from poorer schools, particularly those in the cities with their higher proportion of minorities. The admissions system's responsibility is (1) to ensure the quality of the incoming student population so as to maintain the institution's standards of excellence,

(2) to identify candidates with the competence at least to absorb the educational offering, (3) to minimize freshman year attrition rates in order to maintain the institution's economic viability, and (4) to ensure that public and private employers look to the institution as one providing high quality graduates. With these objectives and their accompanying pressures, and continuing the example above, candidates from the Montgomery County schools would be the safer bets and the safe choice would more often than not be the one made.

Test Scores and Income Level

This committee has been asked rhetorically if it were surprised that test scores and income level are strongly related. The fact is that tests scores and income level are strongly related. It is, however, a simplistic look at the problem. A fuller exploration of the subject would show that families with higher income include more two-parent families, more higher educated families, smaller family size, more value being placed on education, more books in the home during the early formative ages, and better diet for both mother and child in the early years, along with attendance typically at better secondary education institutions. All of these things are related to higher scores and higher performance in school. It also should be clearly understood, especially to those who have raised the question as to whether high school grades would suffice as a sole measure of academic success, that high school grades also show these kinds of relationships.

Test Coaching

H.R. 4949 requires testing agencies to provide statements concerning the extent to which test preparation courses improve test subjects' scores. This is an impossible and meaningless task.

There can be no question that test coaching, particularly admissions test coaching, has become a burgeoning industry. There can be no question, too, that the provisions of the bills having to do with the release of test answer sheets and answers will add to that growth and make the coaches job easier. They no longer will have to resort to circuitous means to get their raw material.

There is still considerable controversy on the extent to which coaching changes scores and more importantly on the real value of the changes, if indeed they do occur.

It is reasonable to assume, however, that coaching for most persons will result in an increase in the score they otherwise would have gotten. It is difficult to measure the magnitude of the difference and the extent to which other variables will affect the differences which will occur. Based on studies conducted on other elements such as the performance of various subgroups on tests containing content more related to their subgroup and the effect of the removal of time limits on speeded tests, it is reasonable to expect that majority enrollees who are already higher scoring will in such courses have a score gain at least as much, perhaps more than minority enrollees. If one considers all of the things related to higher income that were discussed above, it also is reasonable to expect that enrollees in these courses from higher income families will achieve changes at least as high and probably greater than low income enrollees. In short, it is unlikely that establishing funding to ensure that all persons have access to coaching schools will do much more than wildly enrich the coachers. Those who would have scored higher without the coaching will be more likely to take advantage of it and will achieve greater gains. The differences therefore, will not be erased and may well be increased.

There also is a question as to the value of reporting score changes. Score changes are not the true measure of the effectiveness of coaching schools. At this point, there is no reason to assume that score changes brought about by coaching will result in equivalent changes in college grades. Those individuals who are coached will be test smarter, but there is serious question as to whether their mastery of content area has been increased to the point where it will make a difference in college grade achievement. For example, coaching on analogies will improve analogy-solving skill so that those coached will do better on that type item. The question remains as to whether that improved test-taking skill improves performance in college. The solving of analogy problems for some reason is related at least in part to college grade achievement. It is, however, not a direct sample of the performance behavior; that is, getting good college grades. Since the relationship is not clearly direct, a coached ability to solve analogy problems may reduce or eliminate the predictive power of that kind of item for the individual involved. The true test of coaching schools and coaching techniques, therefore, lies in their validation and that validation need only would be satisfied by a substantial showing that the increased performance on the admissions test as a consequence of coaching was also related to increased performance in the college setting.

Destroying The Tests

You have heard repeated assurances that the public release of test items and answers will not destroy or reduce their validity. That advisory is patently false on both counts and the record should show that the profession does not agree. In a telegram sent to Governor Carey prior to the passage of the New York legislation, the American Psychological Association made it abundantly clear that Section 342 of that bill would result in decreased predictive validity of the tests. The Association further saw no purpose being served by the release of test items and answers called for under that Section.

As to assurances that this legislation will not destroy tests, that must be responded to in terms of test type. At one end of the test spectrum is found the SAT, the ACT, and tests commonly described as general aptitude instruments. These tests, being of a more general nature, have a wider body of knowledge from which to draw and the problem of repeated item creation is less acute than with any other test type. The tests which are the primary focus of these bills, therefore, will be most likely to last longest in terms of the release provision. The use of such tests will diminish,however, and costs will go up, as will the use of other tests and less reliable kinds of information. Test administrations will become less frequent; if one has to release test answers every time a test is given, why give it so often to accomodate student and subgroup needs? Research into alternative methods also will diminish. Since new devices will be compromised immediately and copyrights of developers will have been rendered meaningless, the economic incentive will be gone.

At the next point in the testing spectrum, one will find the full array of achievement and knowledge tests. Again, some of these tests will have broader bodies of knowledge from which to draw and new item possibilities will be greater, As a general rule, however, it can be stated that the more specific and therefore more limited the field measured by a knowledge or achievement test, the more likely it is that the creation of new items will become increasingly difficult and that the quality of these tests will diminish as a consequence of less precise measurement. Ultimately, the breach of security provisions of these bills will result in many of these tests' demise as meaningful indicators. Also, if the quality of these instruments does substantially diminish, as indeed it will, what alternatives are suggested by the proposers? All you have heard to date on that subject is euphemisms.

Finally, at the far end of the test spectrum, you will find interest inventories, judgement tests, leadership and creativity measurement instruments, and perhaps most importantly, autobiographical questionnaires. If there is one characteristic which is shared by these instruments, it is that there are no right or wrong answers in the traditional sense. "Right" answers typically are determined on the basis of empirical comparisons across populations.

You have heard repeatedly that traditional admissions tests do not measure the personal characteristics of those taking them. That is true, at least in a direct measurement sense. You have heard, too, that these characteristics include such things as achievement orientation, motivation, judgment, and creativity, and that these characteristics must be measured to obtain the fullest measure of the potential contribution and success of an individual. That also is true. There should be fuller consideration of such elements.

These constructs, or traits as they may be called, perhaps can be measured best by autobiographical information questionnaires. These paper and pencil instruments are based on the simple logic that personal characteristics are predictive and that past experience and achievement predicts future achievement. Biodata instruments contain much of what is found in interviews, only the instruments are standardized--al persons are asked the same questions in the same way--and their record of validity in occupational settings is available and extraordinarily impressive. The "right" answers; that is, those which receive a score are those which are determined by empirical research to be selected significantly more frequently by the successful or unsuccessful, however that is defined. A question concerning high school grades received is a biodata item. The item would call for the respondent to indicate the grades he or she has achieved, or to describe his or her class standing in high school, or both. If higher grades were shown by research to be significantly related to academic or occupational success, then higher grades would be given a positive weight in the scoring system. Biodata items also can be constructed to measure the significance of achievements other than grades, such as participation in athletics, or in social or community activities. Additionally, they can tap motivation, achievement orientation, self concept, and work orientation and values. It is important, too, to note that autobiographical questionnaires seem to be resulting in significantly lower adverse impact than the traditional tests, including the PACE used by the federal government. In short, autobiographical questionnaires show great promise as an alternative or complement to the more traditional tests.

Having said that, however, it also must be pointed out that while adverse impact may be reduced by the use of autobiographical questionnaires, they cannot eliminate what the SAT and the ACT tell us about the educational preparedness of low scorers. For example, a young man in the District of Columbia schools, a valedictorian of his class, was reported to have been denied admission to a university on the basis of his SAT scores. Had those scores been coupled in terms of their use with autobiographical data--his grades, his achievement as a valedictorian, and I am sure the other activities that this young man engaged in, the impact would have been diminished and most likely the young man would have been selected. At the same time there are educational deficiencies which the SAT unmasked. This young man, as is true with hundred of thousands like him, is being cheated. Education costs are increasing at an almost geometric rate, but there is no commensurate increase in the quality of education; in fact it is going down and increasing numbers of students are leaving schools unable to read and unable to compute. As far as other skills are concerned, young girls are coming out of the school systems, particularly the large city schools, with high school diplomas indicating mastery of clerical skills, but when asked to sit down at a typewriter and take a typing test, they manage to produce 20 words a minute, if that. Are these tests wrong?

Returning to the autobiographical information questionnaire's validity and promise in the educational setting, if either H.R. 3564 or H.R. 4949 are passed with provisions requiring release of test questions and answers, then the use of biographical information as an alternative will be permanently foreclosed. Given the nature of the instrument, releasing the right answers destroys it. I would urge you most strongly if you wish a fuller discussion of the possible utility of biodata, to invite experts in the field such as Dr. William Owens from the University of Georgia, where there has been extraordinary use of such a system; invite Dr. Melanie Baehr, from the University of Chicago, who said in a recent symposium that she has been using biodata for years and has found no adverse impact.

Again, however, there will be and can be no question from any source that if either one of these bills are passed with this provision of release in them, biodata will not survive. In fact, its promise will be foreclosed before it can be considered.

The Size of the Task-Conclusion

An ETS staff member is alleged to have said, "We are the gatekeepers." That is an extraordinarily arrogant comment, matched only by its injudiciousness. Nevertheless, we are talking about proposals to legislate how we measure our excellence and any precipitous action by this Committee would be even more injudicious.

As an example of the complexity of what you are being asked to control, the federal agencies involved in civil rights enforcement activities, namely the Departments of Justice and Labor, the Equal Employment Opportunity Commission, and the Office of Personnel Management engaged in the development of what are called the Uniform Guidelines on Employee Selection Procedures. These guidelines have as their stated purpose the provision of advice to the employing community as to how to demonstrate the job-relatedness of their tests. The effort the develop uniform guidelines began in 1972 as a recognition of inconsistencies between standards published by the Office of Federal Contract Compliance, the EEOC, and the then Civil Service Commission. The process took six years and the final document is still the subject of controversy and heated contention. The adequacy of certain of its provisions will ultimately be tested in the courts and there is question as to the extent to which the agencies have comported with professional standards.

As another example, the Advisory Panel of which I was a member accepted no givens and took almost two years to reach its conclusions. Extensive reviews were made of existing data and thirty-seven new research studies were commissioned to provide the answers we sought. Even with that length of time and that amount of probing, we felt that our work could have gone on. I think it important to note, too, that I would expect that the members of that Panel would unanimously oppose the legislative proposals before you.

In summary, this Committee is being asked to produce legislation which is perhaps as far reaching as anything it ever has been asked to do, and you should not be seduced by the blithe assurances that you have received concerning the minimal impact that can be expected from these bills. You are being asked to venture into an extraordinarily complex and critical subject, one which touches most of us at one time or another, and you are being asked to reach conclusions on the subject in a six to eight week time span.

At this point, there is a new law in the State of New York. There is a less onerous law in the State of California. Two states have considered such measures and have rejected them. The National Academy of Sciences also has gathered extensive testimony on ability testing as part of its review of the subject, and a response from the Academy is due to be reported in mid-1980, not an unreasonable time. It would seem at the minimum that this Committee should defer action of any kind on these measures until experience under the state legislation, including its possible revision or repeal, is gained and the National Academy of Sciences report has been published.

We are talking about the techniques we use to measure preparedness and the risks of tampering with them are incredible. What if the tests are right and their critics wrong? Where are the alternatives we can turn to if tests are made less effective or destroyed? Gardner's Excellence also pointed out that, "Anyone attacking the usefulness of the tests must suggest workable alternatives. It has been proven over and over again that the alternative methods of evaluating ability are subject to gross errors and capable of producing grave injustices." In one respect, what you have been hearing is analogous to requests that permission be granted to tear down a building to determine if its architecture is sound. The question must be asked if we tear down the building and we do indeed determine that the architecture was sound, who will rebuild it, what will be left? Again, in the last testimony a member asked what if the tests are right, an incredibly salient question which was met by silence. Finally, the statement that it appears that only the testing companies and publishers oppose these bills is patently false. The opposition is being expressed by the overwhelming majority of our colleges and universities. Opposition already has been stated by the psychological community and these bills are opposed by the business community as well. On balance, this legislation appears to be a child only of those with a hidden agenda, the destruction of all testing as we know it. The American Society for Personnel Administration therefore opposes both H.R. 3564 and H.R. 4949 as ill-conceived, ill-timed and unnecessary. We would urge you to reject or shelve them in their entirety.

Thank you for permitting us to appear before you. We would be happy to respond to any questions which you have.

Mr. **Weiss**. Thank you very much.
Dr. Fleishman?

STATEMENT OF DR. EDWIN FLEISHMAN, PAST PRESIDENT, DIVISION OF INDUSTRIAL AND ORGANIZATIONAL PSYCHOLOGY, AMERICAN PSYCHOLOGICAL ASSOCIATION, PRESIDENT, ADVANCED RESEARCH RESOURCES ORGANIZATION

Dr. **Fleishman**. Mr. Chairman, my name is Edwin Fleishman, and I am happy to have with me Dr. Mary Tenopyr, current president of the American Psychological Association, Division of Industrial and Organizational Psychology. She would be an additional resource to the committee.

I start here as a representative of the American Psychological Association's Division of Industrial and Organizational Psychology, of which I am past president.

I am also past president of the American Psychological Association's Division of Evaluation and Measurement and of its Division of Engineering Psychology, and am the current president of the International Association of Applied Psychology.

I have been a professor at both Yale University and the University of California, and from 1971 through 1976 was the editor of the Journal of Applied Psychology, a primary outlet for the publication of research on test development and validation.

I appreciate the opportunity to address those aspects of the Weiss and Gibbons bills that deal with occupational testing. We support the desire to increase the protection of consumers of testing services; however, we have serious reservations about specific provisions of the two bills.

My testimony focuses on the impact of the bills on occupational testing. I do so only to highlight those issues most directly applicable to tests used in the workplace.

The adverse effects of certain key provisions of the bills upon educational testing have already been amply spelled out by some of my colleagues in earlier testimonies before this subcommittee. Some of these concerns have their counterparts in the use of occupational tests in organizational contexts. I shall elaborate upon them shortly.

Tests are used for a wide range of personnel decisions including those that involve assigning military recruits to different technical training schools, assessing the outcome of some training program, or deciding which job applicants to accept for the limited number of job openings. The use of occupational tests for selection purposes is one that most people are familiar with and one that is expressly envisioned by the langauge in the Gibbons bill, which refers to occupational admissions tests as "any test which is used as part or all of the basis for admitting or denying admission to an individual to any occupation in or affecting interestate commerce." Accordingly, the selection context of occupational testing is my point of reference.

A personnel selection problem is one in which many are called but not all can be chosen; they occur in both industrial and educational settings. In most organizations, tests are used over again on a daily basis with administration to job applicants as they apply. Occupational tests are normally not administered on a regular schedule, as are scholastic aptitude tests. Whether they occur in industrial or educational contexts, selection problems require that decisions be based on relevant information.

That is to say, if Mary is accepted or hired and Tom is not, it is because Mary has a higher probability of succeeding in the school or job than Tom. In the language of the profession, the procedure in question must be validated.

I would like to present a highly telescoped scenario of what test validation involves so that we might have a common framework for the discussion to follow. There are several professionally acceptable ways of validating tests for their job relatedness. One way of demonstrating job relatedness and one that is commonly used for tests of abilities is by means of a criterion-related validation study.

Briefly, the following steps are involved:

One, a job analysis is conducted to determine what sorts of work characteristics are important for competent performance of the job in question.

Two, tests believed to measure the important skills and abilities identified in the job analysis phase are then developed and administered to a group of study subjects or job applicants.

Three, those who are subsequently placed on the job are followed up and measures of job performance are obtained.

Four, at this point in the study, we have collected two kinds of data: Scores on the test and measurements of job performance.

The next step consists of conducting the necessary statistical analyses to determine if and the extent to which performance on the test is correlated with performance on the job.

The process I have described is rigorous, empirical, time consuming, and expensive. It is also pervasively regulated. For example, current Federal testing guidelines embodied in the "Uniform Guidelines on Employee Selection Procedures" require that, whenever technically feasible, the process of validation be performed separately for different race and sex groups for the purpose of determining whether a test that is valid for one group is also and equally valid for other groups.

Test validation is a prominent activity among psychologists. The results of validation studies are written up in technical reports; the most significant ones are published in refereed journals, for example, the "Journal of Applied Psychology." In addition, validation studies are summarized and examined in comprehensive reviews undertaken from time to time to integrate the state of the art or to elaborate upon some theoretical point of view.

The point I make is that the technical reports and professional journals that are the natural outlet for the outcome of validation studies are in the public domain. Consistent with the profession's tradition of sharing the outcome of scientific investigations, they describe in some detail the methods, procedures, and analyses by which it is determined that particular tests are valid for the jobs to which they are geared.

These documents are the most relevant and meaningful sources of information for ascertaining whether occupational tests measure the abilities and work characteristics related to the job. I am asserting that those who would require disclosure of actual tests and test items for the purpose of determining, by visual inspection, whether they are valid are asking for the wrong kind of information.

Since we have been conducting validation studies for some time now and reporting the outcomes of these efforts, what do these studies show? I think the following are most relevant to our deliberations:

One, the evidence is fairly strong that occupational tests tend to show validity for important work behaviors in different work settings. The degree or level of validity is sufficient to render these tests useful in the typical selection situation.

Two, how do tests stack up against other selection devices, such as the interview, letters of recommendation, and indexes of academic achievement, such as grades or rank in class?

The evidence from these studies is fairly strong that tests have a better track record of validities. This finding should not be very surprising when one considers that tests are more standardized in administration and more objective in scoring than the typical subjective and unstructured interview, the glittering generalizations which may characterize letters of recommendation, and the fluctuations in grading standards so common across different schools.

I am suggesting that vis-a-vis these other devices, the typical occupational test of ability has built-in procedural advantages that contribute to its better overall record of validity in the workplace.

Three, the evidence is fairly strong that the relationship between test performance and job performance is linear throughout the entire range of the test score distribution. This means that the higher the test performance, the higher the performance on the job. This means that if for some administrative reason it is decided to set the qualifying or passing score at, let's say, 25, the examinee who scores 30 is better qualified than the examinee who scores 25.

Among those who are considered qualified by virtue of meeting the passing score, some are better qualified than others. Hence, hiring the best qualified job candidate that the labor market will allow is a professionally sound business practice.

Four, the evidence is fairly strong that a test which is valid for one group will also be valid and to the same extent for another group. Use of the same test and the same test standard is appropriate for all race and sex groups. Put another way, if for some reason an organization decided to hire only blacks, it would still be appropriate to use validated tests to select the best qualified from among black applicants.

I conclude this part of my testimony by stressing the fact that occupational testing is pervasively regulated.

First, there are the professional standards governing test development, validation, and use, such as those embodied in the American Psychological Association's standards for educational and psychological tests and the principles for the validation and use of personnel selection procedures developed by APA's Division of Industrial and Organizational Psychology.

Second, there are the uniform guidelines on employee selection procedures adopted by Federal Governmental equal employment opportunity enforcement agencies and given substantial weight by courts in cases where testing is the subject of litigation.

Third, there are the Federal courts with jurisdiction over all aspects of employment discrimination, including occupational testing.

Finally, where occupational tests are used in the movement of bargained-for employees, the grievance and arbitration machinery contained in most union-management contracts provides another effective safeguard. Those who would argue that occupational testing is not regulated seriously underestimate the force of the foregoing sources of regulation. Occupational licensing will probably soon be subject to the same type of regulation. In fact, occupational testing is so pervasively regulated today that psychologists have become, in the words of one lawyer, "familiar forensic fixtures" in employment discrimination lawsuits.

Another effect of regulation, besides insuring test validity and proper use, is to require specific tests for individual jobs. Thus, those employers who test give many different tests for different jobs and it is a rare employer who gives one general test like the scholastic aptitude test. This fact becomes important in assessing the consequences of the proposed legislation.

At this juncture, while recognizing the desire to enhance the quality of educational and occupational testing services in the direction that safeguards the interests of the many consumers of those services, I am compelled to point out the adverse consequences that flow from the application of certain provisions to occupational testing. Two of these are particularly troublesome.

Literally applied to testing in the workplace, section 6(c) of the Gibbons bill reduces the likelihood of hiring the best qualified. The law would require qualification levels to be set at the minimal level of competency. Selection above that level would have to be random.

The idea of being satisfied with minimal competency in an age when national productivity is on the decline, competition for American products and services is increasing, and industrial jobs are becoming technologically more complex is a source of puzzlement to us.

Furthermore, the imposition of minimal standards is contrary to the finding I had cited earlier that the relationship between test performance and job performance is linear. Quite simply, there is no professional justification for being satisfied with less than the best.

We see the provisions in both bills which call for the full disclosure of test items as having the following adverse consequences:

One, test disclosure of the magnitude contemplated in both bills will depress the validity of occupational tests. Validity, you will recall, has to do with the concept of job relatedness. If a test is valid, individual differences in test performance are associated with individual differences in job performance.

The use of only tests having validity is now mandated by Federal law and professional standards. When some examinees have access to the test items while others do not, the former would get an undue advantage on the test which would not be reflected in increased ability to do the job.

Put another way, of two individuals with the same probability of succeeding on the job, that person would be selected who has had prior access to the test items and/or could afford to go to test coaching schools. This situation would operate to the disadvantage of minority job applicants and exacerbate further group disparities in the human condition.

Two, test disclosure would lead most employers to abandon the use of standardized occupational tests of ability. The need to fill openings as they occur and the requirement by regulation to have different tests for different jobs makes the test regeneration process almost impossible. Thus, that which is the most objective component in the selection process would be abandoned in favor of such alternatives as the interview, reference checks, letters of recommendation, and the like.

As I have indicated earlier, the validity of these alternative selection procedures is questionable. Furthermore, they do not enjoy the procedural advantages of standardization and objectivity that are the hallmark of occupational testing and they lend themselves to all kinds of conscious bias.

Three, full disclosure of test items would drastically shorten the useful life of most tests and require the continuous generation and evaluation of new replacement items. The consequent research effort and dollar investments required to maintain the quality of occupational testing under these conditions could become prohibitive. In addition, the expense and effort required to develop new items would discourage research on new and innovative methods for selecting the best qualified individuals for various occupations.

Issues dealing with test disclosure are not new. You may be familiar with the *Detroit Edison* v. *NLRB* case. In a situation involving two legitimate but competing interests—the union's request for access to test and test-related materials in order to properly represent its membership; the interest of management for the continued security of the same materials—the Supreme Court upheld the company's refusal to yield the requested test materials to the union. No less of an authority than the U.S. Supreme Court has considered the arguments over test disclosure issues and passed its judgment on the matter. These points have been adequately covered before.

To conclude, the adverse consequences I have mentioned are more than inconveniences. Their cumulative and long-term effect would be to eliminate the most valid and objective component in the selection process.

Without a clear and overriding mandate, indicated, for example, by documented instances of flagrant test abuse or evidence that mechanisms already in place for requesting testing and for handing test feedback are not working, I urge extreme caution in adopting legislation based on the vague and speculative claim that the bills would increase the accountability of test publishers.

Whatever else the so-called sunshine in testing bills may accomplish, they should not generate such side effects as would endanger what may well be an important national resource for identifying the best qualified people to fill increasingly complicated jobs.

Mr. WEISS. Thank you very much.

Dr. Hall?

STATEMENT OF DR. JUDY E. HALL, PRESIDENT-ELECT, AMERICAN ASSOCIATION OF STATE PSYCHOLOGY BOARDS

Dr. HALL. I am Judy Hall, executive secretary of the New York State Education Department, State Board for Psychology; I am also president elect of the American Association of State Psychology Board, and it is in that capacity that I am here today.

I want to comment on the Gibbons bill, H.R. 3564, and assume that some of the sections in it which are sufficiently vague may permit disclosure at some point, so I am going to include some comments on disclosure on the effect of licensing examinations in my comments.

It is my pleasure to represent the American Association of State Psychology Boards, an association of 54 boards authorized by State or provincial legislatures to certify/license psychologists.

The purposes of this association are:

To facilitate communication among member boards;

To sponsor collaboration between and among member boards in developing compatible standards and cooperative procedures;

To represent the member boards in serving the public interest in matters of psychological services to other psychology organizations, to legislative, judicial, regulatory, and executive governmental bodies;

To aid member boards in fulfilling statutory, professional, public, and ethical obligations.

AASPB was founded in 1961 at the instigation and support of the American Psychological Association, which recognized the need for an independent body to enhance the interstate mobility of qualified psychologists.

It was to that end that a standardized examination was developed in 1964 which could be used by all statutorily recognized bodies as part of their determination of who is minimally qualified to practice the profession.

The examination provides an index of the candidates' knowledge base of the scientific core and the major fields of psychology at a level that all candidates, regardless of their specialties, might be expected to reach.

The questions are developed and selected to measure the examinees' ability to integrate and apply this knowledge, the capacity to exercise ethical judgments, and the attitudes of the applicants toward the profession.

The examination for professional practice in psychology is specifically designed as a test of knowledge, not applied job skills. From that base of knowledge the practitioner can decide which skills need to be applied or, if necessary, what new procedures, derived from the knowledge base, will solve the problem.

Note that this exam is not designed to be the sole evaluative criterion on which licensing decisions are based. It is expected that it will be supplemented by other means of assessing a candidate's competence to practice psychology as a professional.

The examination committee of AASPB consists of carefully selected, prominent psychologists representing the various specialties in psychology as well as having additional expertise and experience in test construction. They, like State board members and AASPB officers, serve without compensation and at great financial loss to themselves for the days that they devote to licensing issues. The decisions related to test construction, validation, setting of passing points, reporting of scores to candidates are all ones reached after many years of experience with the profession of psychology and the state of the art of psychometrics.

Please do not attempt to change our approaches without careful consideration of the potential outcomes. Test construction is our field and we take our jobs very seriously. Let me explain now some of the test construction issues and the progress we have made in providing solutions.

The comments will relate to specific provisions of the Gibbons bill, H.R. 3564.

VALIDITY

Validity is not only an end product of an exam, it is a process in the development of an examination.

In addition to using psychologists to develop items for the examination through periodic item drives, panels of experts in the various specialty areas review the items after each item has been checked for accuracy, relevancy and conformance to psychometric principles. These items are then included in the samples sent to the examination committee for the construction of a new form.

At this point at least 15 people have reviewed and approved each item. Involvement of a large representation of practicing psychologists at every stage of test development is for the purpose of insuring that the examination content is relevant to the practice of psychology.

Up until recently there were no funds specific to research on the examination. In spite of that many States accomplished their own studies utilizing approaches which could be characterized as content, concurrent, and construct validity.

More recently, however, the State boards urged AASPB to increase the fee charged to the States for the use of the examination so that moneys could be invested in validation research. At the present time approximately $50,000 is being spent on four content validation research proposals. The effect of the result of this research will be the construction of an even better examination.

SECURITY

The construction of the examination is a very difficult task since items tapping the functions of a psychologist are ones which involve higher mental processes. Items are such that the candidate may be expected to see the underlying principle or dilemma even though he/she may not be familiar with the specific setting.

These items are limited in number. Therefore, it is imperative that provisions be made to make the exam secure.

At the present time the examination is given twice a year on national testing dates. Once the examination has been completed by the candidate, the candidate does not see the examination again. Should the candidate fail the examination, the specific form can be released but only to the State board who is expected by contract and State mandate to maintain security.

Any State board can request a rescoring of the examination by hand as well as a copy of the answer sheet.

The reasons for maintaining tight security relate to the fact that should candidates have ample opportunity to memorize questions and should those questions be reused, the State board cannot guarantee to the public that those examinees are minimally competent to practice the profession. The greater public good is secured by denying individuals easy access to forms of the examination.

If items were easily constructed and forms were equated without the use of old items, there would be no question about allowing more detailed feedback to failed candidates. But the examination forms have to be constructed with a percentage of these formerly used items. Professional decisions are not easily made, nor are items easily constructed which capture the essence of a professional's thought processes.

If, for some reason, the boards were forced to disclose the licensing examination in psychology, there would be no question that the construction of new forms would proceed at a snail's pace.

The immediate result would be fewer administrations and fewer candidates entering the profession. The end result would be a significant drop in the numbers of providers of health care and in consultants to business and industry.

SUBSCORES

No subscores exist on the national licensing examination since the items which are the best items are not easily classified nor are they mutually exclusive of any other categorical placement.

With the length of the examination kept purposely at no more than 200 items, subscores would not be reliable nor valid.

PASS/FAIL POINT

The individual State board is legally responsible for determining the passing point. The use of two procedures—normative approach and absolute standard—reflect the fact that psychologists cannot decide which is a better approach for the establishment of a cutoff.

Each mechanism has its own set of strengths and weaknesses. The choice of one method over another will not prevent abuses. For the past several years a rationale for the establishment of a passing point has been discussed by AASPB.

A new approach being explored is a methodology using Bayesian statistics. By combining obtained score data with other and preexisting data on candidate competence—for example, academic degree, years of experience and national normative test data—the cutoff could be established after estimating the loss associated with making an incorrect decision.

The articulation of a rationale is one of the priorities of AASPB in the immediate future. We psychologists are uniquely qualified to develop these procedures in the psychometric area. But simplifying a complex matter by mandating one method over the other will only complicate matters.

FEEDBACK TO CANDIDATES

Score information, relative standing, standard area of measurement on the exam, the cutoff point used in that State are all now available to the candidates in each State.

At any time a brochure on the examination for professional practice in psychology can be obtained from the State board or from AASPB. In addition, a brochure outlining the reasons for not being admitted to the licensing examination is available. This brochure is designed for those who fail to meet the statutory requirements of licensure.

CONCLUSION

What I have tried to summarize is the voluntary efforts of those concerned with licensing/certification of psychologists. These have occurred in anticipation of sunset review and prior to the formalization of the EEOC guidelines.

These are in agreement with the American Psychological Association's standards for educational and psychological tests and are in response to meeting the larger public need of endorsing only those who demonstrate themselves to be qualified to practice the profession of psychology.

I would like to recommend that no legislation be passed that would freeze the development of these voluntary procedures but that, instead, we encourage research on testing such as the validity studies we are presently funding or such as the National Academy of Science's study on ability testing which is soon to be released.

Just as AASPB, representing the licensure boards, is independent of the professional organization, American Psychological Association, which in turn represents members of the profession, the examination must be independent of the academic institutions.

These independent checks on the system are important factors in the development of quality professional care. Since the majority of psychologists are health care providers, we are tampering with the potential services to our constituents when we no longer make secure the examination.

Chairman **Perkins**. Thank you very much.

Let me indicate that without objection the brochures and pamphlets that you mentioned in the testimony will be included in the record.

[The information referred to follows:]

INFORMATION FOR CANDIDATES

EXAMINATION FOR PROFESSIONAL PRACTICE
IN PSYCHOLOGY

AMERICAN ASSOCIATION OF STATE PSYCHOLOGY BOARDS

THE PURPOSES OF THE EXAMINATION PROGRAM

The Examination for Professional Practice in Psychology (EPPP) of the American Association of State Psychology Boards (AASPB) is provided for the purpose of helping the state and provincial boards of examiners in psychology to evaluate the knowledge of applicants for licensure and certification. There exists a wide diversity of educational preparation among the approximately 4,000 applicants who seek licensure and certification each year. This diversity is one reason that the state and provincial boards, acting collectively through the AASPB, construct and annually utilize a common, standardized yardstick (the EPPP).

The resources of the whole profession and of the Professional Examination Service (PES) are utilized in the development and continued improvement of the test on a national basis -- resources which are not available to an individual board. The EPPP is only one part of the overall evaluation procedure used by state and provincial boards. The Association expects that the EPPP will be supplemented by other assessment techniques to determine candidates' competence to practice psychology as a profession.

The EPPP is intended to evaluate the knowledge of the candidate with the equivalent of a doctorate and one or two years of experience by sampling the knowledge basic to the practice of psychology at a level that candidates, regardless of their specialty, may be expected to have attained; and by assessing ability to integrate this knowledge and to exercise ethical judgments. Most applicants with the required academic preparation should be able to pass the test with little or no additional study or other preparation. PES does not have copies of past examinations which it can send to applicants, nor does it have a list of recommended books or other material for use in preparation for the examination. However, a candidate can compile such a list by considering the entry-level knowledge necessary for professional practice in the areas covered by the test content outline.

TEST CONSTRUCTION

The EPPP is developed by the Examination Committee of the AASPB in cooperation with the Professional Examination Service, which provides examination services in many health-related fields. The development process is designed to maximize the content validity of the examination for its intended use.

The Examination Committee decides upon the content areas to be covered in the EPPP. Questions (items) are solicited from psychologists representing all areas of the practice of psychology in the United States and Canada. Item writers are provided with instructions describing the kind of items used in the examination. Since the beginning of the program in 1964, more than 500 psychologists have contributed questions for the examination.

The intent is to develop items which measure important knowledge which is directly related to the practice of psychology. Emphasis is on the knowledge that has widespread application and significance in the solution of problems encountered by psychologists in all specialty areas.

The questions are screened and edited by the Examination Committee and by test specialists on the PES staff to ensure subject-matter accuracy and conformance to psychometric principles. They are then sent to panels of three experts for independent subject-matter review. The reviewers' comments are used by subject-matter consultants and test editors in the final review of each item. Items which are approved through these steps are included in a file of items from which the Examination Committee constructs each new form of the test. At the conclusion of this process, at least 15 people have reviewed and approved each item.

In the development of a new form of the test, the Committee reviews each item for continuing relevance and for psychometric performance as shown by item analysis. Each state or provincial board may review the test before electing to use it, and the opinions of the boards on test content are sought and used by the Examination Committee in the development of subsequent forms of the test.

TEST CONTENT AND ADMINISTRATION

The examination consists of objective, multiple-choice questions covering the basic areas of psychology. Each form of the examination contains approximately 200 items. The areas below do not represent separate parts of the examination but are provided as a general outline of the content of the examination as a whole.

General Outline of the
Examination for Professional Practice in Psychology

-History and Systems
-Physiological and Comparative Psychology
-Sensation and Perception
-Learning and Memory
-Information Processing, Thinking, and Language
-Motivation and Emotion
-Individual Differences, Intelligence, and Developmental Psychology
-Personality and Social Psychology
-Behavior Disorders
-Program Evaluation
-Research Design and Interpretation
-Statistics and Quantititive Methods
-Test Theory, Development, and Validation

General Outline (cont'd)

- -Knowledge of Professional Affairs
- -Professional Conduct and Ethics
- -Clinical Psychology
- -Counseling Psychology
- -Educational and School Psychology
- -Industrial/Organizational Psychology
- -Behavior Modification
- -Biofeedback

Samples of the items which appear in the examination are given in the back of this brochure. Each item has four alternative responses, only one of which is correct. The score is the total number of correct responses. There is no penalty for guessing, and it is to the candidate's advantage to answer each item even when unsure of the correct response. No credit is given for items for which more than one response is selected.

The time allowance is sufficient for most candidates to complete the examination.

The examination is administered by the state and provincial boards and the answer sheets are scored by PES which reports the scores, along with relevant normative data, to the administering board. Candidates are identified in the scoring process only by number. Each board sets the standard for passing in its respective state or province, and reports the results to the candidates. All procedures and decisions with regard to licensure are the responsibility of the individual boards.

VALIDATION OF THE EXAMINATION

Content validity is maximized by the extensive test-development process. In order that the validity of the examinations may be continuously assessed, the Examination Committee conducts an ongoing survey of the characteristics of the candidate population. Along with the examination materials supplied to state and provincial boards is a questionnaire which candidates are requested to complete and which is designed to collect data on demographic, educational and experiential background. Candidates are identified by number. The questionnaire data are compared with the results of the examination — total score, pass/fail, and whether or not a license was granted. The results of the survey are reported periodically to AASPB and the state and provincial boards. AASPB has formulated and is currently implementing further content and external-criterion validation studies.

THE INTERSTATE REPORTING SERVICE

An Interstate Reporting Service has been established to facilitate the endorsement of certificates and licenses among states and provinces. The Service maintains a permanent record of examination scores for those candidates who choose to register. At a candidate's request, the Service will report the score, accompanied by normative data such as to assure equitable comparison of scores over time and across test forms, to the board of another state or province in which the candidate seeks licensure or certification. Candidates may request registration of their scores with the Service at the time of examination or any time thereafter. The Interstate Reporting Service applies only to the national written test. Other requirements for licensure in the states or provinces are determined by the individual boards.

The Interstate Reporting Service is maintained by PES. To register scores on the EPPP, candidates should write to the Professional Examination Service, 475 Riverside Drive, New York, New York 10027, and should include the date and place of examination and their unique identification number of the examination. Forms for this purpose are available at the time the examination is administered or can be obtained from PES at the above address. Candidates who do not remember their identification number should inquire of the board in the state or province where they were licensed or certified. Once a score has been registered, the candidate may at any time thereafter request that this score be reported to another state or provincial board. The fee for initial registration of the score is $25.00; the fee for reporting the score to another board is $5.00 for each report made.

FURTHER INFORMATION

For information about procedures and requirements for licensure and scheduling of examinations, a candidate should write to the Secretary of the Psychology Licensing or Certification Board in the state or province in which licensure or certification is being sought. A listing of addresses of state and provincial boards is published annually in the June issue of the American Psychologist.

SAMPLE QUESTIONS

A. According to the Standards for Providers of Psychological Services, a patient's clinical record may be available, even without the client's consent, to members of:

1. The referring agency associated with the client's treatment.
2. The client's immediate family.
3. The professional staff who certify that the record will be used for teaching purposes.
4. The professional staff associated with the client's treatment.

B. In developing a weighted application blank which is to be used in hiring sales agents for your firm, the first step should be:

1. Look over the application blank now in use, assign weights to items you consider significant, then proceed to use the scores you can now get from the blank in appraising job applicants.
2. Administer a preliminary aptitude test battery to all of the sales agents, and examine the scores.
3. Make up a list of productive sales agents, then study these agents to see what characteristics they have.
4. Make up a list of the best sales agents and a list of the poorest agents, and study the two groups to see in what ways they differ.

C. The approach to behavior problems which attempts to reduce anxiety through the elicitation of a response incompatible with anxiety is known as:

1. Operant shaping.
2. Reciprocal inhibition.
3. Pavlovian conditioning.
4. Conditioned reflex therapy.

D. A test of adjustment is administered to 100 subjects and those scoring in the bottom 10% are selected for intensive therapy. Following the conclusion of therapy, the test is readministered and an improvement in scores is noted. Such an improvement in test performance would probably be expected even without therapy because:

1. There has been a lapse of time between the first and second administrations.
2. Such tests are notably unreliable, particulary when based on small samples.
3. Regression of scores toward the mean is to be expected as a purely chance phenomenon.
4. The range for which the test was designed has been restricted by the method of sampling.

E. Recent research suggests that sleep during dream periods tends to be accompanied by characteristic physiological bodily changes. Two of these are:

1. Low voltage, fast EEG waves and REMs.
2. Alpha EEG waves and shallow breathing.
3. Kappa EEG waves and minute generalized movements.
4. High voltage EEG waves and vasodilation.

Answers: A. 4 B. 4 C. 2 D. 3 E. 1

EXAMINATION COMMITTEE

The Examination Committee is appointed by the Executive Committee of AASPB. Its members are chosen for their outstanding abilities and achievements in their specialties and provide representation from major areas of the field. The present members of this Committee are:

George W. Albee, Ph.D.
University of Vermont
Burlington, Vermont

Paul J. Hoffman, Ph.D.
Menlo Park, California

Roger A. Myers, Ph.D.
Teachers College
Columbia University
New York, New York

Wilbur K. Rigby, Ph.D.
Community Mental Health Center
Mission, Kansas

Virginia Staudt Sexton, Ph.D.
St. John's University
Jamaica, New York

Raymond R. Shrader, Ph.D.
University of Tennessee
Knoxville, Tennessee

William G. Steiner
Bradley University
Peoria, Illinois

Michael Wertheimer, Ph.D., Chairperson
University of Colorado
Boulder, Colorado

PES Staff Members:

Craig G. Schoon, Ph.D.
President

I. Leon Smith, Ph.D.
Division Director

Shanna Richman, Ph.D.
Program Associate, Psychology

Professional Examination Service
475 Riverside Drive
New York, New York 10027

Calendar of Common Testing Dates

The Executive Committee of AASPB approved the following schedule for the administration of the EPPP for the six-year period 1979 through 1984:

1979	April 20 October 19
1980	April 11 October 10
1981	April 10 October 16
1982	April 16 October 8
1983	April 8 October 14
1984	April 13 October 26

Mr. WEISS. Mr. Erwin, in the course of your oral testimony, although I could not find it in the written part, you referred at one point to—and I am quoting—a hidden agenda.

I wonder what you mean by that and whether you think that the authors of the bill or the committee, or others at the congressional level, are engaged in attempting to do something other than what is set forth clearly in the legislation?

Mr. ERWIN. I think that many of the supporters of the bill see it as leading to a demise of testing as we presently know it. I don't think the drafters of the bill had that in mind, but I do think that many of the supporters, in terms of the rationale for their support, see it as leading to the decline in testing and possibly the elimination of testing.

That is what I mean when I refer to a hidden agenda.

Mr. WEISS. You also referred somewhere, and again perhaps you will correct me, but as I recollect it and wrote it down, you refer to legislation as urging more restrictive use of tests.

Now, if in fact I have that correct, where in the legislation dealing with postsecondary admissions testing, H.R. 4949, does it suggest more restrictive use of tests?

Mr. ERWIN. I think the legislation's provisions that have to do with the release of test answers and test items is a restricting element in terms of the use of tests, certainly not directly, but certainly indirectly at the minimum.

I think the testimony repeatedly has talked in terms of the effect on tests and specifically on certain kinds of tests, and I would include restricting the use of tests.

I would include within that phrase achievement tests where the items will get increasingly difficult to write and increasingly poor. So that the test's validity will indeed diminish. With some instruments, it will lead to their immediate removal, instruments such as biodata where because of the inherent nature of tests, those tests will stop.

Their performance will be foreclosed before we have an opportunity to even study them. The profession is to a substantial degree coming around to the notion that perhaps autobiographical information can perhaps be an alternative.

There was an example I cited about the District of Columbia youngster who was a valedictorian of his class and was turned down by an institution because of his SAT scores. Biodata would have measured, and there would have been weights associated with the fact that he was a valedictorian and I am sure he engaged in a lot of other activities that would be of a positive nature.

This legislation, in requiring that answers be provided to questions, would foreclose the use of biodata before it ever got off the ground.

I think in the Federal sector where perhaps 25 blacks pass for each white that passes, if biodata were added to the Federal selection system, that there would be a substantial reduction in the adverse impact that that system now has on minority classes.

It is not going to eliminate the situation that the SAT and the ACT tests are telling us are there, but at least it will bring into the measurement scheme a lot of what people have been saying before this committee, we ought to be taking into account personal characteristics over and above how an individual does on an SAT score or on a PACE exam. There are ways to do it, namely, biodata is one of the major ways.

The University of Chicago has been working with it for quite a few years, Melanie Bare. She made a statement recently that in all the years she has been using it she has less, if any, adverse impact from the use of this system. The legislation, what it should consist of, will foreclose the promise of those kinds of systems because biodata, if you provide the answers to the questions, you have destroyed the system by doing so.

That, in essence, is what I meant by restricting the use of tests inherent in the legislation.

Mr. WEISS. I think you may have wanted your statistics to be in reverse. It is 25 whites for every black. I think you said 25 blacks to every white.

Mr. ERWIN. It is reversed, excuse me.

Mr. WEISS. Again, without getting into substance of the tests, the legislation does not deal with that at all. I am just curious as to whether it is your view that biodata is not currently being used by most admissions offices?

Mr. ERWIN. That is correct, in its standardized forms, in the form professionals would describe it, biodata or personal characteristics are used in the admissions system.

People do take account of origins, experiences or achievements of one kind or another. It is done in a relatively subjective mixed way.

Mr. WEISS. You are still talking about admission to post-secondary schools, right?

Mr. ERWIN. That is right.

Biodata, as referred to in the professional sense, is a pencil and paper instrument, the items of which tap such things as prior achievement, motivation to achieve, judgment, work orientation, work values, self-concept. It does this in a standardized scorable way. It eliminates any need to make subjective judgments. It is an empirical system. It has been used in industry for a number of years, decades, in fact.

It has been shown to be one of the most valid predictors of occupational success in industry. It has not been used except in rare circumstances in the academic admissions setting.

If you wish to find out more about its use in academic admissions, Dr. William Owens of the University of Georgia has been experimenting with standardized biodata questionnaires as part of the University of Georgia procedures and has a substantial body of data on the extent to which biodata predicts success in academic settings.

Mr. **Weiss**. I am really puzzled by that whole area because it would have occurred to me, and certainly the testimony that we have received up to this point from everybody, just about, in the testing industry, from the colleges and universities, professional schools, indicated to us in some uncertain terms that the standardized tests are only a very small portion of the basis on which they evaluate applicant's admission and that they take into account the applicant's scholastic record, biographical record, achievements, extracurricular activities, et cetera.

It sort of reminds me about the old story that he didn't know they were speaking grammar until somebody told them.

I don't understand why or how this standardization of biodata, if you will, differs from that which admissions people have been telling us they use in any event.

Dr. Casteen, I wonder if you could comment on how individual biographical information is or is not used currently?

Dr. **Casteen**. Mr. Weiss, I don't hear well. Perhaps I missed the last part of the question.

Mr. **Weiss**. I am not speaking too well today.

I wonder if you would comment on the extent to which the individual biographical background of the applicant is used by you and other admissions people, how it is used for the applicant's application?

Dr. **Casteen**. We have three sources of that. One is the application form, where the student responds in longhand or typewritten form to the questions we propose, questions about their goals, academic experiences, academic expectations, sometimes information about possible careers.

I use one question that says, describe yourself, and gives the student a blank space.

My purpose for that question differs from what Mr. Erwin is describing as the biodata form because I am looking for information and evidence that the student can write and think effectively.

So I use the material on a relatively subjective fashion, based partly on my own experience as an English teacher. I am looking for things that go into other areas of evaluation rather than those described in the biodata research.

Other than that, I am looking for things that might mean different things to different people. I have committee members who will read that information.

The other two resources are the school report form, a document filed on behalf of the student by a teacher, counselor or some other school official, and finally, in some cases, letters of recommendation, especially letters of recommendation that come from responsible adults who have watched the student grow and who can describe in discursive terms a subjective measurement of how well the student has dealt with obstacles, what the student does in the face of resistence, how the student is to articulate a goal and possibly accomplish it through a combination of academic and extra-academic activities.

The purposes of the information I collect that are used commonly in the more selective colleges and universities differ in basic concept from those pursued in industry and the kind of instrument that Mr. Erwin is describing.

We are not interested in scoring at the moment, in scoring a student's biographical experience. We are interested instead in seeing how well the student thinks in describing it, in seeing how effectively the student can relate that experience to his academic and personal intentions, and finally, in seeing whether the student writes a reasonable example of standard American English prose.

Mr. **Weiss**. A number of you have commented on the effects of coaching and I think that there was some testimony which related to how, by providing the answers as well as the questions, you are in essence making the job of the coaching institutions easier, and that likely it is going to be those who have the wherewithal to take the coaching courses who will be taking them.

My question is, Mr. Erwin, you referred earlier to my entering into the record the item from the New York Times a few weeks ago indicating the prevelance of coaching institutions right now.

Isn't it a fact that coaching really goes on at a great pace at this point, and what happens currently is that although it may be more difficult for the coaching institutions to get the answers to questions, what they are primarily concerned with, that in fact they get it nonetheless.

Certainly you have indicated that there seems to be a correlation of improvement among those who take the coaching courses, that in fact what you would be achieving by opening it up is not just to have coaching schools burgeoning but that you would have people who currently do not have access to coaching institutions in fact be able to compete on a more equal basis with those who currently can afford to take the coaching courses?

Mr. **Erwin**. That assumes that people who have the answer sheet by themselves on their lonely basis can improve by themselves without the coaching. I don't think that is a reasonable assumption.

Mr. **Weiss**. No, it doesn't assume that. It assumes, for example, that the NAACP which testified before us in support of this legislation may decide to undertake a coaching course on their own to provide them to students.

Mr. **Erwin**. There would still remain the basic question, does coaching make a difference in terms of subsequent performance? That question has not been answered by any of those testifying. That question has to be answered.

Mr. Weiss. You have indicated in your testimony that you think in fact it does improve scores.

Mr. Erwin. Test scores, yes, I do. I think so. But I think also that the question still remains, do those who get coached perform at an equally higher level when they get into an academic setting? That question is unanswered.

Mr. Weiss. But that really has no relevance because the admission is dependent on the score, really, to whatever extent it is dependent on the score rather than on the basic fundamental underlying educational component quality.

Dr. Casteen?

Dr. Casteen. Mr. Weiss, may I add a description of three experiments that I have undertaken having to do with test scores and the relationship between test scores and academic training?

I am not interested specifically in coaching, but for several years I have been asking students whose records are generally sound but whose SAT scores include subscores that indicate reading deficiencies and secondly, also students whose reading deficiencies can also document from other standardized testing that appears in the school record, I have been asking those students on a random basis to undertake some kind of reading improvement program, sometimes a commercial program or a program available at the school aimed at increasing speed or rapidity of reading, improving general reading comprehension and teaching students how to deal critically with what they have read, and I have laid down those three stipulations to be pursued in the programs.

Almost inevitably, students who have taken that kind of reading improvement aimed not at the tests but dealing with the skills that the tests rewarded and also the skills we rewarded in teaching students at the undergraduate level, I have found almost inevitably that test scores improve dramatically for students who take that kind of remediation in the 12th grade.

I have for several years asked Virginia residents whose records are generally sound but whose high school programs indicate they have not taken the most rigorous program available to them until the senior year to wait until after the senior year of high school, which for many of these kids is their first experience with five academic courses in the course subjects, and then retake the SAT.

I have published an experiment this year about the outcome of that improvement because we found the mean improvement of those students exceeded the FTC report on coaching.

Third, we run each summer a summer preparatory program for students with nonacademic backgrounds who are Virginia residents, mostly black students. It has nothing to do with standardized tests. But it deals in a basic intensive fashion for 6 weeks with reading, writing, and arithmetic.

I will be glad to furnish you the outcome of that program, also, because at the end of the summer we established a SAT for those students and found score improvements that go far beyond what the coaching companies have been able to demonstrate and furthermore indicate the direct link between the kinds of academic skills these youngsters are not acquiring in school and the kinds of skills they must have to do well on the tests and succeed in the rigorous undergraduate program.

It is demonstrable that students who learn to read effectively, write, and subtract and multiply effectively get better test scores and higher grades when they get into college. That, I think, is the core of the problem in coaching.

Mr. **Buchanan**. Mr. Chairman, I might say at the outset, that while I appreciate the contributions made by the members of this panel overall, there are two that are of special interest to me.

I did graduate work at the University of Virginia which is very ably represented this morning by Dr. Casteen; and Dr. Hall came to her present position from the university that is my largest constituency, the University of Alabama in Birmingham. So, I have also a provincial interest in the testimony.

I am one of the members of this committee who is undecided in this whole area, because, while I think something is the matter, I am very much afraid it's mostly something the matter with the educational system at the elementary and secondary levels rather than with the testing.

I would like to know for certain, if it is fact, so that we can take more decisive action, so we can take steps to correct the problems. Some of the problems may be with the testing, for all I know.

As I understand Chairman Weiss' bill, H.R. 4949, the idea is to make possible the research that that would further evaluate those tests, and to make sure that there are no restrictions on such research, as well as to let individuals have more freedom pertaining to the information as to how they performed on the tests.

Now, that may or may not be an accurate explanation of what you are trying to do, but it is as I understand it.

I guess there are several things about which I have questions, but let me ask rather generally and ask you to respond.

First, is this an area where Congress needs to act? Does there need to be Federal action in this area? The State of New York has acted although I know that the primary sponsor, State Senator Kenneth LaValle of New York, has indicated that he is going to look at certain aspects of his law during the next legislature to see if there are areas of revision that should be made in the New York law.

So, first of all, is this an area where we need congressional action?

Second, if so, is this legislation specifically the action we need to take? Dr. Hall has indicated that she is less than happy with some of the provisions of the Gibbons bill and the impact that they might have. Others have spoken more to the perceived consequences of the Weiss bill.

Is there specific action we can take, and is there action we can take, perhaps a little different from either proposal, that might nail down whether or not there is bias in testing?

Third, if Congress should act, should we act now? Is there anything that would indicate that we need to take a harder look before we act, or is the situation such that we need immediate action. Those are all general questions, but I would appreciate your responses to them.

Dr. **Fleishman**. Well, I think the position that the professionals that I represent are taking is that we don't need to act now, that there are a number of existing regulatory forces that are taking into account some of the concerns that the committee has expressed.

Also, there are some developments now going on, such as the National Academy of Sciences study which is getting complete, and the passage of legislation in other States which are yet to be evaluated. It would seem to us that it's premature to enact legislation at this time unless these reports and actions can be properly evaluated, so that is how I would respond to your question at this time.

The third point is that there are side effects of current proposals which would in fact reduce the validity of existing procedures, and would also not solve the problem of reducing bias and discrimination.

Mr. **Erwin**. I would agree in terms of should we act now, that it's premature. I pointed out earlier the advisory panel on score decline took 2 years to reach its conclusions, and even then we were not satisfied. We commissioned 37 separate studies of the subject. The Federal agencies involved in establishing the uniform guidelines on employee selection procedures took 6 years to reach their conclusions.

There is law; several States have rejected the similar law. There is the National Academy of Sciences study which is now ongoing. All of these things argue in favor of a longer observation period.

I don't think that what is being proposed is the action that you need to take. I think the longer evaluation period will help bring that to a little bit sharper focus. I think it is abundantly clear from the testimony that you have heard, not just from the testing companies but from the professionals in general, that one of the major problem elements in the legislation is the call for release of test items and test answers.

This in fact will depress validity and is one of the major obstacles to a fuller discussion or meeting of the minds.

Now, do you need to act? Act is a broad word. The question is, Do you need to legislate?

I don't think you do at this point in time. But I think the role of Congress ought to include a little bit of leverage in place of legislation. It may very well be that some meaningful discussions can begin. If we share common objectives, that is to say, what is it that can be done to get the data out into the open so that more people can study the data, and the conclusions can be drawn quicker, more fully, and the concern about a few bodies having the keys to the kingdom, if you will, anything that the Congress can do in its leverage role, not its legislative role.

I don't think it would be a waste of this committee's time at the staff level to be talking to the testing companies and saying, let's talk about if we agree on the need to get data into the open so that people can study it, responsible professionals can see it in its rawest form and not a summary form, how can we do that without bending your noses too far out of joint.

I think a responsible discussion can take place at that level, and we may reach agreement. That is possible.

I don't think that the Congress needs to just sit still. I think you have incredible influence in terms of the leverage factor, and that that leverage factor ought to be used.

Mr. BUCHANAN. Thank you very much.

Does anyone else have a response to those concerns?

Yes, sir?

Dr. CASTEEN. Briefly it may be useful to you to know that many of the concerns that the committee members have expressed and that have turned up in testimony on both sides of this question to this committee are also the common concerns of those of us interested in teaching and moving students effectively through the educational system.

That does not answer the question as to whether legislation is necessary. For example, one of my activities as dean at Virginia is sitting on one of the councils by which the college boards seeks to receive opinions, judgments, responses from its college constituencies.

That council for 3 years encouraged the college board in an ongoing project to release a full copy of the SAT with relatively detailed explanatory data that could be directed to students and used in helping students prepare for the test, and that discussion predated by a good bit congressional interest in this issue of disclosing the contents of not every test but in this case a typical and authentic test.

The book was published, as you may know, last year and goes to students who take the SAT's on a regular basis. That indicates the fact that some of what Mr. Weiss has been concerned about has been the subject of concern in the professional community, but it also indicates the extent to which a community that initially developed testing because of a clear need in our educational system and because of the peculiarity of the American education system.

We do not have a dictated national curriculum. Every organized industrialized nation can indeed measure what it intended by education because the State dictates what it needs in the class.

Given the origins in the testing system and our educational program in general, the fact that the community has been able to respond to perceived needs when they come indicates initially that the private system works.

On the other hand, it's going to be the case in the course of history, we are involved on a daily basis with students and their credentials and will miss points that will be obvious because we are too close to trees and too far from the forest.

Congressional scrutiny is a healthy and necessary phenomenon. In my earlier appeal for a serious national look at the whole question of why schooling does not accomplish its stated purposes, and I think that should be the proper business of Congress.

Mr. BUCHANAN. Thank you very much.

I must say this has been a particularly helpful panel, and I appreciate your testimony.

Just one further thing, Mr. Chairman, and I appreciate your patience.

Mr. Erwin, you especially, I think, have been thinking the unthinkable and speaking the unspeakable here this morning.

It does seem to me that we need to do a good deal more of this. You and Dr. Casteen have both talked about some of the, I am afraid, very real problems that we are facing, and of what is happening in education in the United States at the elementary and secondary level.

I am very much afraid that we have not had enough frank, honest facing up to the realities of the problems. I really think that the standardized tests may be perceived as something of a test for the teachers as well as for the students, and if that is true, then both the teachers, and the students have been flunking the test and I think that is not and could not be a happy thing for a professional group.

I would say we may discover that the teachers are not the heart of the problem after all. We have laid unspeakable burdens of varous kinds on some teachers in recent years. They, in many places, are an underpaid, underappreciated group of people who have had too little voice in policy administration. Yet, I very much fear that if we look hard enough and face the complexities of the problem with enough initiative and enterprise, we may find there are some very real, very severe problems in elementary and secondary education in the United States.

One of my great concerns is that if you don't have clearly valid standardized tests there to measure how we are doing, whether it deals with the problems of an ethnic group or a socioeconomic group, or with the problems or people in general in the United States, I very much fear we will lose something valuable by which we can help or which can help us analyze and try to deal with the problems as they are.

That is a really long speech rather than a question; but if anyone would care to respond, I would be pleased.

Mr. Erwin. I think that on one count there is a question that says: "How is your wife?" and the response is, "Compared to what?"

That has to do with norm reference testing as opposed to criterion referenced testing. The provision in Mr. Gibbons' bill that has to do with a prohibition against comparing people in terms of their relative standing, if that kind of external measure, how do we stand against the rest of the community, other communities, or the Nation, and if we are not permitted to make those kinds of relative evaluations, one school system as opposed to another, then the parents will have no measure of how well their children are being educated because there will be no comparisons available.

They will be truly the captive of the system. There are some questions about how much do you really want to find out. We know that there is an illness. There is just no question about it. All of the measures that we are saying, and it's not just the SAT which is, incidentally, declining, something that causes us some concern.

We have seen it in the math skill reports that come out of the National Assessment Association on Educational Progress; and we did see it in Dr. Jensen's writing again in the Washington Post of October 5, the Washington Star of October 4, where there is a report that the academic quality of young teachers is taking a nose dive. That is only one study, but to the extent that SAT scores are declining generally, they are declining just as well for people who are going to be prepared to be teachers as they are for people who are going to be chemists or engineers or doctors or whatever.

There are some very difficult bullets that we have got to start biting, and that has to do with the quality of education for our children and what happens before we get to the schools.

You are right, teachers in elementary schools have an extraordinary burden in the school systems generally. They are becoming the caretakers or expected to be the caretakers. That is not the way the system should work. There has to be some preparation before schooling, and that is taking place to a lesser degree than it ought to.

We ought to do more about the effects of diet on different segments of the population, on us all. We are all the same. Deprivation is very fair. It's much fairer than anything else in the society. It affects all of us equally, and there are some who are more deprived than others, not only in terms of the education they receive but what happens before they begin to take their education.

Those things have got to be addressed, and we have got to start addressing them. I think we need a truth in education bill or at least start talking about a truth in education bill rather than just a truth in testing bill. We are still dealing with the thermometers, and that is inappropriate. The thermometers surely have to be looked at, but we can't ignore what the thermometers are telling us, and to a substantial degree we are.

Mr. **Buchanan**. Thank you very much.

Any other response?

Thank you, Mr. Chairman.

Mr. **Weiss**. Mr. Erdahl?

Mr. **Erdahl**. Thank you, Mr. Chairman. I apologize for not being here for the panel's presentation; but I will read your testimony.

Mr. Buchanan asked the question I was wondering about. I think it's well that this bill is before us, whether it passes or not, but I think that we have seen the highlighting of a concern, whether it's a concern over the results of the testing or a concern over what brings those results about. I guess I would echo what the last witness said in a different word perhaps, that maybe we are pursuing the wrong villain. The villain could be a combination of things, whether it's deficiencies in our educational system, in the parents—and I am a parent with several children in school—whether we as parents have failed, whether our children have spent too much time in front of the television, or whatever.

I would hope that with the benefit of this testimony and other statements that we have received both pro and con that we could come out of our deliberations with something that truly is good for not primarily the teacher or the system or the Government, but the student. That is the main concern we ought to share as members of this committee and as Members of Congress.

Thank you, Mr. Chairman.

Mr. **Weiss**. Mr. Erwin, in the course of your testimony you referred, I think I am almost quoting, to the arrogance of the Educational Testing Service spokesperson who said that their job, they said: "We are the gatekeepers."

I gather from that that you do not consider that the appropriate role of the testing industries to keep people out of institutions of higher learning.

Mr. **Erwin**. I had a feeling that you would pick up on that statement and had considered leaving it out but thought I would leave it in.

That is one staff member; I thought it was. If it were true, it was reported to me secondhand, and if it were true it was arrogant to say that. I think it was inappropriate and not really the function of testing as we see it. It was also injudicious, to say the least.

I don't think the mentality exists, and we were exposed for 2 years during the course of the advisory panel's convenings to a great degree to ETS staff and generally found them to be extraordinarily helpful and very, very sensitive to their responsibility in terms of the role that they had to play as professionals. They are a very professional group of people, a very professional organization.

I don't think they see themselves as the gatekeepers in terms of keeping people out. They have taken substantial strides and from a point of view that tests to a substantial degree are the thing that get people in.

If you do agree that people evaluate school systems, they know the relative quality of school systems. Admissions personnel, that is their function to know how well school systems do in terms of educating the people that come from them.

The tests very often help to reach out within school systems where you know the quality of education is generally poorer. The tests help to reach out, to pull in, if you will, those people who are otherwise not pulled in, so I don't think they see themselves as gatekeepers in the sense of selecting out at all. I really do not.

I think it was a poor remark to make, if it was indeed made, and I was using it to go on to the broader point that I was making in the testimony.

Mr. **Weiss**. I appreciate your utilizing it, because to me it sort of highlights part of Dr. Casteen's testimony in response a little bit ago to a question from Mr. Buchanan, about when he referred to the fact that people in the admissions field themselves and people in the field of testing generally have raised some of the very questions and concerns that we have been raising in suggesting legislation, H.R. 4949.

I have been impressed by the testimony from professionals who are not part of the testing industry who have spent their lives, some for decades, some for 5, 10, 15 years, in the field who have given us examples where, in spite of what the Educational Testing Service, for example, pretends to make available raw materials to qualify professionals where some of them have had to go into court to gain access to materials of that kind, so the legislation that I am sponsoring does not have it as its aim the destruction of testing.

What I am concerned about is, in fact, fair testing. I think that fair testing can only come about if professionals in the field have a chance to really look at the data, the raw materials, and to critique, because I don't think that the byproduct, the end result of testing is only of concern to the testing industry.

I think that it's a national concern, and I think that when we have so many millions of our people having their entire lives, never mind careers, dependent on some of these tests, there ought to be as broad an evaluation as to what those tests really test.

Are they really testing aptitudes? Are they testing, as you suggest, Dr. Casteen, reading and writing capacity, and so I have tried to frame this legislation to be very, very strict in adhering only to disclosure requirements and not of trying to dictate to anyone what ought or ought not to be involved in the making up of those tests and how they are to be interpreted by anyone.

I very much appreciate your testimony this morning.

Before we close, Dr. Tenopyr, you have been sitting there very patiently, and I want to ask you if you have any comment on anything that has been said?

Please, I want to give you the opportunity to say anything and not walk away and say nothing.

Dr. TENOPYR. I would like to reiterate, business and industry are already very carefully regulated, and you should take the Uniform Guidelines on Employee Selection Procedures and enter them into the record. You will find them very interesting, and we are talking about people who can't read and write.

These are written at about the 23d reading level, but we are quite heavily regulated. We are doing a lot of research. All employers are doing a great deal of research to insure their tests are valid, and we did so before the regulations and doing even more so now, and we are trying to take action in other ways and social programs to take care of the inequities that have resulted from the problems of inferior education and deprived backgrounds.

That is all I want to say.

Mr. WEISS. Thank you very much.

Does anyone else have any final words to say before we conclude the hearing?

If not, thank you very much.

Mr. BUCHANAN. I would like to ask that the studies to which Dr. Casteen referred be made a part of the record.

Mr. WEISS. Without objection.

[The information referred to follows:]

The 1979 class includes 81 (±5) Virginians offered admission conditioned on retaking the SAT. Of this group, 21% show no gain, 43% show gain in both tests, 28% show gain only in the verbal test, and 7% show gain only on the math test. Mean gain for the entire group on both tests is 66 points. For those showing gain on the verbal test, the mean gain is 53 points; for those showing gain on the math test, the mean gain is 57 points. Insofar as the standard error of measurement for the SAT is V31 and M33 and the standard error of the difference is V44 and M47, these gains are all well above the statistical model. In other words, the retake experiment works in that it gives students significantly or meaningfully higher scores on both tests.

The typical student in this group is a white female from a Virginia suburban public school. Her high school class rank is high; her academic program contains four solids per year with no more than two A.P. or honors courses; her standardized test record shows reading problems dating from the eighth or ninth grade, and no effort at early remediation. She was 17 on February 1 and 18 on June 1. Her latest SAT scores lift her out of the group of entering students who most closely resemble students on our probation/suspension lists in the past four years. Our estimate is that her new credentials predict first-year work in the range of 2.5 ± .2 and graduation on time with an average of 3.0 ± .2.

45 Summer Prep students, all with histories of difficulty in taking standardized tests, retook the SAT on July 21 under the joint sponsorship of the Office of Admissions and the Office of Afro-American Affairs. Lawrence A. Groves supervised the test administration. 11 (24.4%) showed no improvement on either test. 8 (17.7%) improved on the verbal test, but not on the math test. 13 (29.8%) improved on the math test, but not on the verbal test. 13 (29.8%) improved on both tests. Among those who improved on the verbal test (46.6% of the whole group), the mean

improvement was 42.86 points. Among those who improved on the math test (57.7% of the whole group), the mean improvement was 50 points. A tentative conclusion: the Summer Prep experience, which includes intensive work in the core academic disciplines, but no direct tutoring on the SAT or similar tests, appears to produce improvements in the basic learning skills (reading, verbal reasoning, computation—the skills rewarded by the SAT) that exceed by significant factors the normative score gains reported for all students who retake the test. (E.T.S. reports the standard error of measurement as 31 verbal and 33 math and the standard error of the difference as 44 verbal and 47 math.)

A note of the Test of Standard Written English: 21 students (46.6%) showed improvement, with a mean improvement of 6.56 on the scale of 20–60+. The reported standard error measurement is 3.9; the reported standard error of the difference is 5.4. A tentative conclusion: for students whose command of basic writing skills improves during Summer Prep, the improvement exceeds the improvement reported for all students who improve on retaking the TSWE.

Mr. **Weiss**. If there are no further questions, the committee's appreciation to you again for your patience and perseverance throughout your testimony.

The hearing is adjourned.

[Whereupon, at 12:55 p.m., the Subcommittee on Elementary, Secondary, and Vocational Education of the Committee on Education and Labor adjourned.]

TRUTH IN TESTING ACT OF 1979; THE EDUCATIONAL TESTING ACT OF 1979

THURSDAY, OCTOBER 11, 1979

House of Representatives,
Subcommittee on Elementary, Secondary,
and Vocational Education,
Committee on Education and Labor,
Washington, D.C.

The subcommittee met, pursuant to recess, at 9:30 a.m., in room 2175, Rayburn House Office Building, Hon. Carl D. Perkins (chairman of the subcommittee) presiding.

Members present: Representatives Perkins, Ford, Andrews, Weiss, Kildee, Peyser, Kogovsek, Goodling, Buchanan, Erdahl, and Hinson.

Staff present: John F. Jennings, counsel; Jeffrey W. Brennan, staff intern; and Richard DiEugenio, minority legislative associate.

Chairman **Perkins**. The committee will come to order.

Our first witnesses this morning are Dr. Harold Howe, vice president for education and public policy, the Ford Foundation, and Dr. Gary Keller, dean of the graduate school, Eastern Michigan University.

We will hear from Dr. Howe first before we hear from Dr. Keller and we will withhold your questions until both of these gentlemen testify.

You may proceed in any manner you wish, Dr. Howe.

Without objection, all the prepared statements will be included in the record.

We appreciate seeing you here again this morning.

STATEMENT OF DR. HAROLD HOWE II, VICE PRESIDENT FOR EDUCATION AND PUBLIC POLICY, THE FORD FOUNDATION; DR. GARY KELLER, DEAN, GRADUATE SCHOOL, EASTERN MICHIGAN UNIVERSITY, REPRESENTING THE HISPANIC HIGHER EDUCATION COALITION, ACCOMPANIED BY DR. ALVIN RIVERA, HISPANIC HIGHER EDUCATION COALITION, A PANEL

STATEMENT OF DR. HAROLD HOWE II, VICE PRESIDENT FOR EDUCATION AND PUBLIC POLICY, THE FORD FOUNDATION

Dr. **Howe**. Thank you, Mr. Chairman. It is good to be here and to see you again.

Mr. Chairman, I would like to start by saying that Dr. Kenneth Clark, who had intended to come to these hearings today, was prevented from doing so. He asked me to bring his testimony down and deliver it to the committee which I have done. I have given it to your staff.

I would like to read three very short sentences from his testimony simply to give an overview of what it has to say.

He says:

> I am opposed to the present legislative attempts to deal with the important problem of abuse and misuse of the standardized test results for the following reasons:
>
> First, the present truth in testing legislation is misleading and confuses the public;
>
> Secondly, this legislation will not benefit or protect minority and poor students from being discriminated against in obtaining access to higher education or to jobs; and
>
> Thirdly, the important problem of cultural bias in test results and their interpretation is obscured rather than clarified by this proposed legislation.

The rest of his testimony implies those points and will, I assume, be entered in the record.

[Statement referred to above follows:]

PREPARED STATEMENT OF DR. KENNETH B. CLARK, DISTINGUISHED PROFESSOR EMERITUS, CITY COLLEGE OF THE CITY UNIVERSITY OF NEW YORK

Mr. Chairman: I am opposed to the present legislative attempts to deal with the important problem of the abuse and misuse of standardized test results for the following reasons:

1. The present "Truth in Testing" legislation is misleading and confuses the public.
2. This legislation will not benefit or protect minority and poor students from being discriminated against in obtaining access to higher education or to jobs.
3. The important problem of cultural bias in test results and their interpretation is obscured rather than clarified by this proposed legislation.

In regard to the general problem of misleading and confusing the public, it is clear that the important problems involved in test construction, test validation and reliability cannot be resolved by the unrealistic demand that the "meaning of test scores" be explained to students. The constructive evaluation and the interpretation of tests are highly technical matters which must be dealt with only by ongoing empirical research conducted by those who are trained in this specialty. Specialists familiar with this problem know that there is no such thing as an absolutely perfect test with absolute predictability. But they are also aware of the fact that the abuse of standardized tests cannot be resolved by a simplistic law which confuses this important issue with confuses this important issue with consumer protection problems. To approach this problem from the point of view of legislation will merely give the public the impression that it is being protected when in fact this is not the case. This is not only misleading but may also direct attention from the research which is necessary to increase the validity of these tests.

In regard to the protection of minority and low income students, this law may in fact have the opposite effect. Middle class, middle income families can afford to provide their children with the type of specialized tutoring which they believe will increase their scores. Minority group and low income families will remain victims of the low scores in spite of requirements that testing companies explain the "meaning" of their scores. The gap between the average scores of middle class whites and the average scores of low income and minority students is a function of a complexity of factors which will not be resolved by the requirements of this law. It is conceivable that this gap may be increased by legislation which promises protection which it will be unable to give.

Closely related to the problem of the protection of minorities by a truth in testing law is the persistent and justifiable concern as to whether standardized tests are instruments for the perpetuation of bias and discrimination in our society. Those who are opposed to the professional use of these tests argue that they are culturally biased. The fact of the matter is that there is at present no test which can be culture free. All tests measure aptitudes and achievements which are acquired in a given society or culture.

A truly culture free test would in fact have no predictive value in determining the ability of an individual to function at a given level of performance in a given culture. Standardized tests do not create or, when properly used, reinforce biases in a society; rather, they reflect the biases and inequities which exist in a society. In this regard these tests are valuable diagnostic instruments which when used and interpreted by properly trained professionals provide us with the indications of what the society must do to remedy existing educational, social and economic inequities. Some of us continue to hope that upon the basis of the results of these tests, it will be possible for the society eventually to make the commitment necessary to remedy these inequities. Without these test results, we will not have the indicators of what remains to be done—and what must be done.

THE UNIVERSITY OF THE STATE OF NEW YORK,
THE STATE EDUCATION DEPARTMENT,
New York, N.Y., August 7, 1979.

The EDITOR,
The New York Times,
New York, N.Y.

DEAR SIR: On July 24, 1979, the New York Times uncritically endorsed the "Truth In Testing" legislation, which was signed by Governor Carey. Your editorial stated that this amendment to the education law would take "the mystery out of college testing," and that "students deserve to know how they are being rated and judged."

In spite of the good intentions of the Times editorial, it was misleading. So was Ralph Nader's assertion, in a letter published in The Times on August 3, 1979, that the legislation will require test companies "to explain to students what the scores mean and how they will be reported to schools." And Nader's claim that objections to this legislation reflect only the "corporate style lobbying efforts of testing services" is inaccurate. The Board of Regents and The Commissioner of Education of The State of New York have independently advised the Governor against signing this legislation. It is my contention that this so-called "Truth In Testing" law is a placebo. However laudable its intentions, this law cannot force test companies to explain the meaning of test scores to students; and certainly this law cannot deal with the complex issues of test validity and the role of cultural factors in influencing test results. The construction, evaluation and interpretation of tests are highly technical matters which must be dealt with by on-going research by those who are trained in this speciality. The important problem of the use and abuse of standardized tests cannot be resolved by a simplistic law which confuses this issue with consumer protection problems.

Admission tests measure the abilities developed by individuals over a long period of time, both in and out of school. The value of test results is directly related to the equality of opportunity afforded to the candidates taking a particular test. The assumption that a candidate's examination of a test already taken will somehow improve scores or reduce the chances of abuse is highly questionable. One can further assume that only the more privileged students will avail themselves of the opportunity to examine the test materials. These students will seek private tutoring on the assumption that this will increase their future test scores. Those individuals who cannot afford this privilege will certainly not improve their test scores by merely examining the test questions and scores. This is a deceptive and meaningless exercise.

The New York State version of the "Truth In Testing" law is misleading and confuses the public. It will not benefit minority and poor students. It could, in fact, be detrimental to already-disadvantaged individuals. Appropriate legislation in the general area of the abuse and the premature exclusionary use of standardized tests might be desirable, but such legislation should be the culmination of serious inquiry, rather than a political gesture. It is my firm belief that this is such an important matter that the New York State Legislature should hold a series of hearings with concerned professionals for the purpose of amending this law, which promises much more than it can possibly deliver.

Sincerely,

KENNETH B. CLARK.

Dr. HOWE. My own testimony, Mr. Chairman, has been submitted to you and rather than read it I will simply quickly go over it, picking out some key points, the first of which is that I think this inquiry you are making is an important inquiry and one that I am glad is being made because I do think there are problems in this field of testing and the use of testing and fairness in testing.

But I have to say at the same time that I hope very much that the Congress will not rush into legislation on this subject. I think that the bills before you are not well advised.

Basically, my own view is that this is a subject which, while important, really represents an undesirable entrance of Federal concern about education into a very special field in the schools and colleges.

I draw attention in my testimony to some earlier hearings that you may recall, Mr. Chairman, just before you became chairman of this committee when the then chairman, Rev. Adam Clayton Powell, held some hearings on the history textbooks in the schools.

He was absolutely right. There were serious abuses in those history textbooks, abuses of the record as far as minorities and particularly blacks in America were concerned.

But he did not go on to produce legislation on that subject. He did a great public service by ventilating the issue and that caused some reaction of a useful kind among the publishers.

I suggest that this exercise here could well have a similar effect in the field of testing. I am not trying to say that there are no problems in the testing field, but I would observe that, as Dr. Clark does, that those problems are extremely complex, that the idea of using essentially a consumer protection type legislation to try to get at them will not, I believe, do a great deal of good. In fact, it won't really address the problems.

I think that the major point that is unrecognized in much of the discussion on this subject is that tests really provide for a great deal of fairness in the handing out of educational opportunities in the United States, that by using tests we effectively cancel considerations of race or wealth or political power or other undesirable criteria and really reach for considerations of merit.

Therefore, while there always have been and always will be problems in regard to testing, I don't think we want to attack it when it is one of our great protections from unfairness.

I would call your attention to the fact that in my view at least the testing authorities who work primarily with college admissions testing to enter graduate professions of various kinds have increasingly opened up their activities as well they should and probably they should go further in that direction.

I myself was instrumental some years ago in helping them to think through the business of releasing test scores to students which they did not at one time do. Now they do. The volume and nature of interpretive materials for tests available to students, available to test users, has steadily increased. It ought to increase more.

Again I say, however, that I do not think that legislation by Congress on this subject is the way to bring about that increase.

Finally, Mr. Chairman, I just want to say that I have shared these general views about these matters with a number of educators, among them people well known to you, former Secretary John Gardner, former Secretary of HEW, and Francis Keppel, former U.S. Commissioner of Education.

They concur in these views of welcoming this inquiry, thinking it will do some good, but not being enthusiastic about the idea of congressional legislation on this subject.

That is a brief summary of my testimony, Mr. Chairman. I will be glad to answer questions.

[The prepared statement of Harold Howe follows:]

PREPARED STATEMENT OF HAROLD HOWE II, VICE PRESIDENT FOR EDUCATION AND PUBLIC POLICY, THE FORD FOUNDATION

Mr. Chairman: It is easy to understand the concern of numerous persons about the fairness and the proper use of tests for guiding admission both to education opportunities and to occupations. Tests provide information about the performance of individuals in a great variety of areas, and it is important that both those who take them and those who use them know as much as possible about how they work, how they are administered, how their results are properly used, and what their limitations and shortcomings are. I welcome the inquiry of this committee into the nature and problems of testing. It helps to ventilate issues that should be before the public, and it helps to encourage an atmosphere of openness in an area that has to some extent been characterized by unnecessary confidentiality and secretiveness.

But having said this Mr. Chairman, I have to say at the same time that the two bills before your committee constitute overkill in reaching for solutions to whatever problems exist in the testing field. In my view the enactment of such legislation is unwarranted and unwise. Although I believe that you have done education a service to open up this subject, I strongly recommend that you not rush into detailed national legislation concerning it.

Back in 1966 a former Chairman of the House Education and Labor Committee found himself in a situation somewhat similar to yours today. The Reverend Adam Clayton Powell conducted hearings on the treatment of minorities, particularly Blacks, in the history text books used in American schools. He called the publishers before him, as you have called the test makers, and read them the versions of American history that their texts portrayed, showing in the process that the interpretations of slavery, the Civil War, and the situation of Blacks in America had been tailored to suit the tastes of prevailing opinion in various parts of the country rather than the facts of history. Chairman Powell did a great service in opening up this subject as you have done for the issues of testing.

In the course of those hearings in 1965, I testified, as the then United States Commissioner of Education, that it would be a grave error to legislate on this subject of history textbooks because it would be an unwarranted intrusion by the national government into the affairs of educational institutions. This is exactly my view of the legislation you are considering now. Chairman Powell's hearings did not result in legislation, and I hope yours will not.

This proposed legislation lays a heavy federal hand on delicate matters in both education and employment. It constitutes a significant change in the assumptions under which we operate our schools and colleges in this nation. In the name of "Truth in Testing," a concept that everyone should support, it brings federal authority into student and employee selection processes in great detail. To warrant this sort of intrusion into the affairs of educational institutions and employers would require at the very least clear evidence that fundamental and constitutional rights of individuals are being denied by testing practice. I know of no such evidence. Indeed, it seems to me that test makers and test users are growing progressively more attentive to the problem of minorities, women, and other special groups.

Mr. Chairman, if you will look carefully at the record of two major private, nonprofit organizations concerned with admissions testing in the United States, the College Entrance Examination Board and the Educational Testing Service, you will find that over the years both have moved progressively and voluntarily toward the kind of openness that this legislation advocates. In the early 1950s, students were not even allowed to know their own scores on College Board admissions tests, an unnecessary secrecy that I and others helped to persuade the College Board to abandon. Since then the volume and the nature of explanatory material for test takers and test users has increased many fold. Sample questions, publications on the meaning of scores, and guidance for college admissions officers on the proper use of test scores are widely available today. The American College Testing Service and ETS publish such materials for the tests they provide. So do many commercial test sources.

I am not saying that there are no problems and no abuses in the field of testing, but I do assert that they are not sufficiently serious to warrant legislation which fundamentally alters the nature of institutional freedom for schools and colleges and moves federal authority further toward domination of the education scope. Furthermore, I believe that there are constructive actions you can take that will assist with whatever problems exist without starting down the road that this legislation implies.

The very complexity and detail of the legislation you are considering suggests that the subject of abuse in testing is a complex and difficult one. It is a subject most educators know little about. I strongly suspect that if ETS were to give us educators a test on testing, most of us would flunk it. It is an area where quick legislation designed to satisfy complaints can create turmoil rather than fairness and progress.

For example, it is widely asserted that tests are unfair to minorities. If that is true, the problems involved will not be solved by the legislation you are considering. Dr. Kenneth Clark has noted (New York Times, August 18, 1979) in opposing similar legislation in New York State, "The important problem of the use and abuse of standardized tests cannot be resolved by a simplistic law which confuses this issue with consumer protection problems." Furthermore, there is good reason to believe that the proper use of tests adds a major element of fairness to educational admissions practice in the United States and protects minorities as well as everyone else from arbitrary judgments based on racial discrimination, wealth, political power and other undesirable criteria in the selection process.

I suggest that instead of legislating on these matters you appoint an independent National Commission to review the issues and to report back to you after an extended study of all the allegations that have led to considering the present legislation. Such a group will require both qualities of statesmanship in its members and special skills and knowledge in its staff. While it should have some educators on it, it should not be sponsored by any organization of institutions or educators. It should be truly independent. Its charge should be broad, requiring it to examine the nature and use of testing in deciding about the educational and occupational opportunities of Americans and to make recommendations about problems that need attention from educators, employers, and other parties, including federal, state and local government if the Commission believes that government should be involved. To look into all the matters that have been raised on this subject, the Commission will require extended time and adequate funds.

I have shared the general views presented here with a number of educators and found them supportive. Among them was John Gardner, former Secretary of Health, Education, and Welfare, and Francis Keppel, former United States Commissioner of Education. It was Secretary Gardner and Commissioner Keppel who helped this Committee in the early 1960s to fashion a new role for the federal government in the field of education. They agree with me now that this proposal carries federal regulation of education beyond a reasonable point, that the proposed legislation does not address the issues that need attention, that the dangers inherent in it far outweigh any possible benefits, and that constructive efforts to solve existing problems will be helped by the visibility you have given these matters and could be further helped by responsible inquiries into them.

Finally, Mr. Chairman, I want to recognize that I have not in this testimony commented in detail on the two bills before you. This is mainly because, as noted already, I don't believe that the Congress should pass any legislation in this field. I do want to say, however, that I find many difficulties with various provisions of these bills.

At the present time I am Vice President for Education and Public Policy of the Ford Foundation, but this testimony reflects my personal views rather than those of the Foundation, which does not take positions on matters of this kind. I attach a brief resumé of my experience in education.

HAROLD HOWE II

Born Hartford, Connecticut, August 17, 1918; married Priscilla Foster Lamb, September 1940; three children—Gordon A. Howe, Lt., United States Marine Corps; Mrs. Merrill Howe Leavitt, Denver, Colorado; Mrs. Frederick T. (Catherine Howe) Short, Fairbanks, Alaska. Residence, 340 East 64th Street, New York, New York.

Education

Taft School, Watertown, Connecticut—1936; Yale College—B.A. 1940; Columbia University—M.A. (history) 1947; Harvard Graduate School of Education—Education Studies; and University of Cincinnati—Education Studies.

Honorary degree and awards

LL.D.'s from Princeton, University of St. Louis, Notre Dame, Shaw, Adelphi; Gold medal from New York University; Special Citation from Frank Newman's Committee on Higher Education.

Employment

Darrow School in New Lebanon, New York—History Teacher and Assistant Headmaster, 1940–41.

United States Naval Reserve—Commanding Officer Minesweeper in Atlantic and Pacific Theaters, 1941–46.

Phillips Academy, Andover, Massachusetts—History Teacher, 1947–50.

Punchard High School, Andover, Massachusetts—Principal, 1950–53.

Andover Junior High School, Andover, Massachusetts—Principal, 1950–53.

Walnut Hills High School, Cincinnati, Ohio—Principal, 1953–57.

Newton High School, Newton, Massachusetts—Principal, 1957–60.

Newton Junior College, Newton, Massachusetts—President, 1957–60.

Superintendent of Schools, Scarsdale, New York, 1960–64.

Learning Institute of North Carolina—Director, 1964–65.

United States Commissioner of Education, Washington, D.C., 1965–68.

Program Advisor in Education for the Ford Foundation, New Delhi, India, 1969–71.

Vice President for Education and Research, Ford Foundation, New York, 1971–79.

Vice President for Education and Public Policy, Ford Foundation, N.Y., 1979– .

Extra curricular activities (past)

Mayor's Committee on Juvenile Delinquency, Cincinnati—Chairman.

Committee on Examinations of College Entrance Examination Board—Chairman.

Commission on Humanities of the American Council of Learned Societies.

United States Council for UNESCO.

Board of Trustees of the John F. Kennedy Center for Performing Arts.

Trusteeships: Yale University, Vassar College, College Entrance Examination Board, Taft School.

Consultant to Mayor Edward Koch on the City University of New York, 1978–79.

Extra curricular activities (present)

Trustee of the John Hay Whitney Foundation.

Cleveland Conference.

National Council on Educational Research, National Institute of Education.

Publications

Numerous magazine articles in Saturday Review, Vital Speeches, etc.

Numerous chapters in books on education.

Book of Speeches as United States Commissioner of Education, Picking up the Options—1968

Chairman **Perkins**. We want to thank you for your testimony, Dr. Howe. To my way of thinking you perhaps have been the most outstanding Commissioner of Education in my time here at the Capitol. I know that you served with much distinction during the Johnson administration and you established a great reputation for the building of education throughout the Nation.

I think your testimony will be most helpful to this committee. I hope that you can stay around a few minutes until we hear from the other gentlemen.

Mr. Kogovsek, if you want to introduce the gentleman, we will be delighted to hear from you at this time.

Mr. **Kogovsek**. Thank you, Mr. Chairman. I appreciate the chance to do this.

I want to say how tremendously proud I am to have as a member of the panel this morning Dr. Alvin Rivera from the Hispanic Higher Education Coalition in Washington, D.C. Dr. Rivera is from my home district in southern Colorado.

I want to tell you that his testimony and expertise can prove very valuable and I want to welcome Dr. Rivera here this morning.

STATEMENT OF DR. ALVIN RIVERA, HISPANIC HIGHER EDUCATION COALITION

Dr. RIVERA. Thank you very much, Congressman Kogovsek. I appreciate the opportunity to have this opportunity to present testimony on behalf of H.R. 4949 specifically. The coalition has had an opportunity to find an expert, a person from Eastern Michigan University who truly has expertise and who has written numerous books on this topic.

We would like to ask, Mr. Chairman, if he may provide the testimony in this case.

STATEMENT OF DR. GARY KELLER, DEAN, GRADUATE SCHOOL, EASTERN MICHIGAN UNIVERSITY, REPRESENTING HISPANIC HIGHER EDUCATION COALITION

Dr. KELLER. Mr. Chairman and members of the subcommittee, I will be making several points that go beyond the written version of the testimony we submitted yesterday.

My name is Gary D. Keller and I am testifying as the spokesperson for the Hispanic Higher Education Coalition concerning the topic of educational testing, particularly with reference to H.R. 4949, the Weiss bill.

Today, as you know, I am accompanied by Dr. Alvin Rivera, a former resident of Colorado and an active participant in and supporter of the Hispanic Higher Education Coalition. That group is comprised of representatives of 11 national Hispanic organizations enumerated in the written version of my testimony which are interested in improving educational conditions for their constituencies.

My academic background is as follows: I am the dean of the graduate school, Eastern Michigan University. I have earned a Ph. D. in Spanish and an additional masters degree in psychology. I am the author of 12 books having to do with Hispanic, Spanish and language topics as well.

As a Hispanic educator I have been extensively involved in the creation of tests and assessment instruments primarily in the field of higher education.

The Hispanic Higher Education Coalition enthusiastically supports H.R. 4949 in all but two aspects. Therefore, we wish to make two suggestions concerning this bill:

One: That data be gathered by linguistic and cultural minority groupings and that the testing services be required to report scores and norm scores separately for major discernible linguistic and cultural groups.

Two: That the provisions for disclosure of test questions and answers be deleted.

We support the Weiss bill because it will go a long way toward protecting the Hispanic-American consumer of tests, making the test services accountable for their product, in facilitating more equal access to education for Hispanic Americans, in more accurately defining the appropriate use of specific tests, in correcting the unwarranted mystique surrounding standardized testing, significantly increasing the awareness and sophistication of the general public and the Hispanic communities with respect to the nature of standardized and other tests, and controlling for nonvalidating and nonpredictive distortions of the tests when used with Hispanic populations.

We feel that certain reports with negative implications for standardized tests have not been released. For example, a major study produced by and at the Educational Testing Service entitled "Cultural Bias in Testing: Challenge and Response," prepared by Dr. David Loye of the ETS staff at the request of ETS president Turnbull.

Another example of missing data appears in the report, "California: College-Bound Senior 1979," reported by the admissions testing program of the college board.

In that report, both the relationship between SAT scores and income level and the relationship between ethnicity and income level of those taking the SAT is reported yet unaccountably the relationship between ethnicity and SAT scores is glaringly absent, perhaps because such data might have cast the SAT in a bad light.

The testing agencies need to be compelled to provide the background data and statistics to permit outside, independent and impartial recorders to evaluate the adequacy of the tests.

What little preliminary evidence that there is from outside sources due to the overly protective attitude of the testing agencies is cause for great disquietude.

For example, David White at Berkeley's childhood and government project has reported findings indicating a negative correlation between medical exam scores and performance in actual clinical situations, thereby casting doubt on the validity of those tests as predictors.

Prof. Benish Hoffman's well-known study has demonstrated how the tests use ambiguities and misleading language in order to make their questions inappropriately difficult.

We need to have research conducted to determine the effect of such artificial test language on minority linguistic and cultural groups. We urge that the questions be considered and empirically resolved about a possible interaction effect between artificially created, misleading English test language and the linguistic parameters of the U.S. Hispanic communities.

We need to answer the following question: How much of that standardized test purports to measure in the scholastic aptitudes of Hispanic students, and how much is merely a subterraneous index of the purely linguistic peculiarities of various Hispanic communities?

In this regard we note that apparently because the LSAT was a poor indicator of academic success in the law schools on the island of Puerto Rico, about 10 years ago the Educational Testing Service established a Spanish version of its graduate record examination, Prueba de Aptidud Para Estudios Graduados, PAEG, appropriation for measuring aptitudes for the study of law.

Currently, students seeking entrance to law schools in Puerto Rico are required to take both the English language LSAT and the Spanish language law test section of the PAEG, a test which was independently developed for Spanish-speaking students and which in no way is a translation of the LSAT.

Apparently bilinguals fluent in both English and Spanish who take this test have widely varying scores on the English LSAT and the Spanish PAEG. We need to carefully integrate this phenomenon which on the face of it points to language and culture as very strong determining factors in tests which are purported to measure aptitudes such as for law.

We need to investigate the extent to which testing agencies have either confused or fostered in the general public a confusion between the prognostic value of tests and their diagnostic value.

Prognostic is the process of predicting outcomes. Diagnostic is the process of identifying causes. Tests identified as prognostic function to predict future student performance and have been interpreted as such.

Using aptitude tests for diagnostic purposes is, in our judgment, a test abuse which H.R. 4949 can help control by compelling the testing agencies to specific for what purpose a test has been devised.

Even the late Dr. Oscar Burras, publisher of the Mental Measurements Yearbook and winner of the ETS award for distinguished service to measurement, declared in the Educational Researcher in 1977 that the information available to permit an adequate assessment to be made of test makers secure tests is quite unsatisfactory. I quote him, "quite unsatisfactory."

The reason for this, it appears to us, is that not enough material is made available for adequate review. We liken this situation to having the fox guarding the hen house data.

In sum, the Hispanic Higher Education Coalition endorses the general goals of the truth-in-testing movement, most of the specific goals of H.R. 4949, and most of the provisions by which this bill proposes to attain these goals.

The first aspect of the proposed Weiss bill for which we are suggesting modifications is one of the methods by which the bill attempts to monitor the need to insure equal access.

The bill requires that each test agency shall provide "a comparison of the average score and percentiles of test subjects by major income groups."

In addition, the bill requires that:

> Within 1 year of the effective date of this Act, the Commissioner shall report to Congress concerning the relationship between the test scores of test subjects and income, race, sex, ethnics, and handicapped status.

We propose that H.R. 4949 not only require the aforementioned disclosures which we enthusiastically endorse, but in addition require affirmative action to counter linguistic and cultural biases in the content by means of appropriate modifications in the norming procedures of tests.

The testing services should not be merely required to provide information concerning the average score according to income, but should also be required to report scores by major discernible linguistic and cultural groups and to scale the test scores within these groups.

In other words, we propose that scoring be based on the achievement of a student using his or her language and culture groups as the norm or the control. There is already a well-established precedent within the graduate record examination and other studied tests for this procedure.

In the GRE, for example, the verbal score of any given woman is converted to a percentile score based on group scores of women. Men are given a percentile score based on male performance. This procedure candidly admits that men as a group are disadvantaged compared to women in the verbal component of the GRE and that normalized scores increase their chance to get into graduate school.

This precedent needs to be extended to all of the pertinent standardized tests that affect identifiable language minorities such as the Hispanic American communities. Only in this way can we remove the biases of our standardized tests, if not from the content, then from the scoring procedures and the interpretation of the results.

We wish to emphasize both the purpose and the practicality of the suggestion that we offer to have test scores normed according to identifiable linguistic minority groups.

Hispanic Americans should be normed separately, not in accordance with any philosophy of educational separatism, but on the contrary, so that Hispanic students can be permitted access to the prestigious American institutions of higher education in the numbers that their capabilities warrant.

The norming procedures that we suggest would control for biases in the testing content that hold back significant numbers of Hispanic youngsters from equality of educational opportunity and access. We want more of our children in the mainstream of American education, and that is why we propose these corrective norming procedures.

In further support of our position, it should be noted that there is at least one law in force which establishes the requirement for what we are suggesting.

In section 121a.530(b), Protection in Evaluation Procedures, of Public Law 94-142, Education of Handicapped Children, it is stated that testing and evalution material and procedures used for the purpose of evaluation and placement of handicapped children must be selected and administered so as not to be racially or culturally discriminatory.

The problem has been to enforce this requirement. We believe that the procedure that we are suggesting, requiring of the test agencies that they normalize test scores for identifiable linguistic minorities, would bring us one step closer to meeting the legal requirement of Public Law 94-142, and the intentions of H.R. 4949.

The second aspect of H.R. 4949 about which the Hispanic Higher Education Coalition has reservations involves those portions of section 5 which require the disclosure of all test forms, including all questions on the test, together with the correct answers.

The creation of culture fair test items is a difficult task to perform in the establishment of standardized tests. Questions which have proven reliability with respect to identifiable linguistic minorities need to be carefully husbanded. If they were to be revealed, we don't think that they could be effectively substituted.

Therefore, the coalition is concerned that the good intentions of the Weiss bill might be hindered by the actual disclosure provision so that instead of the purported goal of the legislation to insure equality of educational access, an undesired side effect would occur: A deleterious weakening of the validity of standardized tests with respect to controls for linguistic and cultural bias.

Instead of providing general test disclosure, it is more in the interest of educators concerned with equality of educational access that the test services be required to provide timely, periodic evidence of the nondiscriminatory nature of their tests, both with respect to the scaling and scoring procedures and with respect to the actual content of the questions themselves.

The Hispanic Higher Education Coalition wishes to close this testimony with one last observation. The intent of our suggestions is not to put roadblocks in the way of the testing agencies with a view to ultimately destroying the testing industry. Far from it, we see important benefits to be derived from objective tests, including standardized tests.

For example, Hispanic educators have been involved with bilingual education for the last 20 years and longer. Most Hispanic educators readily acknowledge the extremely important contribution that the development of assessment procedures has made to bilingual education.

Hispanic educators are usually leery of assessment procedures, and rightly so, because of their potential for bias. And yet, in point of fact, bilingual education as we know it today would not have been possible without the massive, often court ordered, as in the *Aspira* consent decree, development of assessment procedures to determine the language competencies of nonnative English-speaking children.

Contemporary bilingual education is irrevocably tied to the good services of assessment procedures in determining placement of children within bilingual classrooms. That is all to the good, for bilingual educators often have many sad experiences to recount in witnessing the illegitimate subjective placement of minority youngsters in special education classrooms, often for the mentally retarded or learning disabled, simply because of their lack of experience with the English language.

Bias among educational personnel is just as real as content bias in tests. Disclosure of test items in the area of bilingual education will serve no useful purpose.

Indeed, to cite one example, it is generally acknowledged in the city of New York that the language assessment battery, the test used to implement the *Aspira* consent decree, is in urgent need of additional, secure versions because a result of the current policy to readminister it over several grades causes children to become "test wise."

Because of repeated exposure to the test, children receive a higher paper score than their language competency warrants and therefore are removed from bilingual education classrooms.

In conclusion, we fully support the movement toward more accountability in testing. We endorse most of the provisions of the Weiss bill. We think that the bill should be modified and strengthened, however, in the particulars that we have described above in order to better attain its provisions for equality of educational access for linguistic minority students.

It is because the Hispanic educator probably more than any other minority group professional has been privy to the worst, most biased aspects of the testing process, as well as its potential for accurately assessing competencies and determining placement, that these suggestions are entered in the spirit of a throughgoing reform of a service that needs to regain public trust in order to continue to be productive and beneficial to society.

I thank you very much for your attention and I stand ready to respond to any questions.

[The prepared statement of Dr. Gary Keller follows:]

PREPARED STATEMENT OF DR. GARY D. KELLER, FOR THE HISPANIC HIGHER EDUCATION COALITION

Mr. Chairman, members of the Subcommittee: My name is Gary D. Keller, and I am testifying as the spokesperson for the Hispanic Higher Education Coalition concerning the topic of educational testing, particularly with reference to H.R. 4949 (the Weiss Bill).

The Hispanic Higher Education Coalition (HHEC) is comprised of individuals representing 11 national Hispanic organizations interested in improving educational conditions for their constituencies. The HHEC has as its central purpose to increase participation of Hispanic populations in higher education in the nation. The 11 national organizations which comprise the HHEC are: ASPIRA of America; El Congreso Nacional de Asuntos Colegiales; League of United Latin American Citizens; Mexican American Legal Defense and Educational Fund; Mexican American Women's National Association; National Association for Equal Education Opportunities; National Council de La Raza; National IMAGE, Inc.; Puerto Rican Legal Defense and Education Fund, Inc.; Secretariat for Hispanic Affairs, U.S. Catholic Conference; and Spanish American League Against Discrimination.

Support for most of the provisions of H.R. 4949

The Coalition supports H.R. 4949 in all but two of its aspects. We wish to make two suggestions concerning this bill:

(1) That data be gathered by linguistic and cultural minority groupings and that these data be normed by such groupings; and

(2) That the provisions for disclosure of test items be deleted.

Hispanic communities have long been victimized, we believe, above and beyond any other identifiable population, by certain kinds of tests, including the principal standardized assessment instruments such as the SAT, GRE, LSAT, MCAT and others, because of the well-recognized linguistic and cultural biases against Hispanic Americans that those tests have built into them. The testing services have not done enough to counter or control for these non-validating, non-predictive distortions of the test scores when applied to Hispanic American populations, and we believe that the proposed H.R. 4949 as well as at least one additional feature in H.R. 3564 represent a first and long-needed step in improving the situation of our community. Our remarks will primarily focus on H.R. 4949, although we recognize the need for greater public accountability in occupational testing as required by H.R. 3564.

The Weiss bill will go a long way in protecting the Hispanic American consumer of tests, in making the test services accountable for their product, in facilitating more equal access to education for Hispanic Americans, in more accurately defining the appropriate use of specific tests, in correcting the unwarranted mystique surrounding standardized testing, and in significantly increasing the awareness and sophistication of the general public and the Hispanic communities with respect to the nature of standardized and other tests.

We fell that H.R. 4949 provides legal requirements that will encourage more openness on the part of the testing services. These requirements have become essential because of the current climate of mistrust within large sectors of the public—including the Hispanic American sector—with respect to the services now provided by the testing agencies. At this time we are very concerned about the charges that various supporters of truth-in-testing legislation have made that certain reports with negative implications for standardized tests have not been reported, for example, a major internal study produced by and at the Educational Testing Service, entitled, "Cultural Bias in Testing: Challenge and Response," prepared by Dr. David Loye of the ETS staff at the request of ETS president Turnbull. The seriousness and substance of charges such as these have led to a critical situation.

The necessary trust that is essential for the proper use of standardized tests has eroded to such a degree in the general public, and within the Hispanic communities, that federal action is required. We feel that the test subjects and postsecondary education institutions will welcome the provisions of the bill requiring the disclosure of information (reports, statistics, and other information) and will restore the public confidence in standardized tests.

In sum, the Hispanic Higher Education Coalition endorses the general goals of the truth-in-testing movement, most of the specific goals of H.R. 4949, and most of the provisions by which this bill proposes to attain these goals.

Need to strengthen the antidiscrimination provisions of H.R. 4949

The first aspect of the proposed Weiss bill for which we are suggesting modifications is one of the methods by which the bill attempts to monitor the "continuous need to ensure equal access for all Americans to educational opportunities of a high quality." [Section 2.(a)(2).] The bill requires that each test agency shall provide "a comparison of the average score and percentiles of test subjects by major income groups." [Section 3.(a)(B).] We believe that in order to better attain equality of educational opportunity, the requirements of the Weiss bill can and should be strengthened in the following manner:

We propose that H.R. 4949 not only require disclosure but in addition require affirmative action to counter linguistic and cultural biases in the content by means of appropriate modifications in the norming procedures of tests. The testing services should not be merely required to provide information concerning the average score according to income, but should also be required to report scores by major discernible linguistic and cultural groups and to scale the test scores within these groups. In other words, we propose that scoring be based on the achievement of a student using his or her language and culture groups as the norm or the control. As John Garcia ("Intelligence Testing: Quotients, Quotas, and Quackery," in Joe L. Martinez, Jr., Chicano Psychology, N.Y.: Academic Press, 1977, p. 207) has noted, there is already a well-established precedent within the Graduate Record Examination for this procedure:

> * * * in the GRE * * * for example, the verbal score of any given woman is converted to a percentile score based on group scores of women. Men are given a percentile score based on male performance. Since men, as a group, score lower on the verbal GRE, each man is assigned a higher percentile score than any woman who does as well as he on the test, and he receives a higher percentile score than many women who do better than he does. (The converse is true of the quantitative component of the GRE). This procedure candidly admits that men, as a group are 'disadvantaged' compared to women, in the verbal component of the GRE and that normalized scores increase their chance to get into graduate school.

This precedent needs to be extended to all of the pertinent standardized tests that affect identifiable language minorities such as the Hispanic American communities. Only in this way can we remove the essential Anglocentric bias of our standardized tests, if not from the content, then from the scoring procedures and the interpretation of the results.

Lest it be argued that the number of Hispanic students who take standardized tests is too small a population to adequately perform the norming procedure that we are suggesting, we observe that according to Table 10 of California: College-Bound Seniors, 1979, published by the Admissions Testing Program of the College Board, in the state of California alone, 7,287 Mexican-Americans and 311 Puerto Rican students took the SAT test. This student population base is more than ample to provide separate norms for Hispanic students.

The Coalition believes that the grouping categories for both disclosure and norming requirements should also include "linguistic and cultural minorities." It can be clearly established that standardized tests are biased for both income groups and for linguistic and cultural minorities and we believe that the degree of bias is more acute for linguistic minority groups. We wish to cite statistical evidence of such biases:

a. In the document, California: College-Bound Seniors, 1979, produced by the Admissions Testing Program of the College Board, Table 11 breaks down mean and median incomes as a function of SAT average. The pertinent data is as follows:

	Below $350	$350-399	$400-449	$450-499	$500-549	$550-599	$600-649	$650 or over
Mean income	$19,500	$25,000	$27,100	$28,700	$30,200	$30,900	$32,400	$33,800
Median income	15,400	20,500	22,100	23,700	25,000	25,500	27,200	28,800

There is a clear, direct, manifest, correlation between income level and SAT scores. The more the income level of the parents, the better the SAT score for all income groups.

b. In Table 10 of the same report, Annual Parental Income by Ethnic Group, the relationship between ethnic origin and income level is reported:

	Mexican-American	White
Mean income	$16,200	$30,500
Median income	13,600	25,100

A comparison of the statistics reported in a. and b. reveals that the mean and median income of Mexican-Americans is below the smallest income category reported (below 350 on the SAT) while white students average in the 500–600 range on the SAT. By comparing the two tables (Table 10 and Table 11 in the report), we find what would appear to be an interaction effect between income and linguistic minority origin. Mexican-Americans do poorly on the SAT because they have low incomes. The combination of factors interact so that the bias effect is made extremely acute for Hispanics.

Apropos of the above analysis, the following comment is in order: Why was it necessary to make such indirect comparisons? Why didn't the Admissions Testing Program of the College Board give us a table which directly compared SAT scores in California with ethnic group status? Perhaps because such data cast the SAT in a very negative light. Hopefully, the HR 4949 provisions will require the clear and systematic presentation of such data, as well as the norming provisions to control for the biases which such data point out.

Just as men and women are normed differently in verbal and quantitative scores, respectively, we suggest the same process be used for linguistic minority groups. We wish to emphasize both the purpose and the practicality of the suggestion that we offer to have test scores normed according to identifiable linguistic minority groups. Hispanic Americans should be normed separately, not in accordance with any philosophy of educational separatism, but on the contrary, so that Hispanic students can be permitted access to the prestigious American institutions of higher education in the numbers that their capabilities warrant. The norming procedures that we suggest would control for biases in the testing content that hold back significant numbers of Hispanic youngsters from equality of educational opportunity and access. We want more of our children in the mainstream of American education, and that is why we propose these corrective norming procedures.

Moreover, let no testing agency claim that it is impossible or impractical to norm in the fashion that we have suggested. In addition to the precedents that we have described, Jane R. Mercer, one of the best-known and most respected assessment experts in the United States, has developed, in collaboration with a number of colleagues, an assessment instrument called: System of Multicultural Pluralistic Assessment. This instrument, published by one of the largest test organizations. The Psychological Corporation (a division of Harcourt Brace Jovanovich), among other procedures, evaluates a child's performance relative to others from the same sociocultural background, thereby taking into account the cultural specificity of tests and the distance between the student's sociocultural space and the culture of the school.

Modeling H.R. 4949 on Public Law 94–142 (education of handicapped children)

In further support of our position, it should be noted that there is at least one law in force which establishes the requirement for what we are suggesting. In section 121a.530(b) (Protection in Evaluation Procedures) of Public Law 94–142 (Education of Handicapped Children), it is stated that "testing and evaluation material and procedures used for the purpose of evaluation and placement of handicapped children must be selected and administered so as not to be racially or culturally discriminatory."

The problem has been to enforce this requirement. We believe that the procedure that we are suggesting, requiring of the test agencies that they normalize test scores for identifiable linguistic minorities, would bring us one step closer to meeting the legal requirement of Public Law 94–142, and the intentions of H.R. 4949.

Concerns about the disclosure of test questions and answers

The second aspect of H.R. 4949 with which the Hispanic Higher Education Coalition has reservations is those portions of Section 5 which require the disclosure of all test forms, including all questions on the test, together with the correct answers.

It is the contention of the testing agencies that systematic disclosure of test questions and test answers in itself will lead to a general deterioration of test quality because it will make pre-testing very difficult and repeated use in future tests impossible. Moreover, its restrictive effect on the testing agencies' efforts to both isolate and reject culturally biased test items and to include those items which do have high correlations not only with the mainstream population, but with all identifiable linguistic and cultural minority populations, will be very greatly hindered, further promoting bias in the content of tests which have been revealed to be biased in the first place. In other words, it is the contention of the testing agencies that the proposed requirements we have alluded to in Section 5 of H.R. 4949 will further increase the Anglocentrism of standardized tests and cause the collapse of the ongoing partial progress that the testing agencies and assessment researchers have made over the last 20 years, to create "culture fair" tests.

To the extent that the testing agencies can provide convincing proof of the validity of their contentions, we are in concurrence with their argument that the test items and test answer disclosure provisions in H.R. 4949 should not be enacted. Our position is that in addition to the suggestions that we have offered to modify scoring and morning of standardized tests, the test services should be made more accountable for reducing the current biases of standardized test item content.

The provisions of the Weiss bill ought to be strengthened along the lines of section 121a.530 of Public Law 94-142 which we have cited earlier, so that testing agencies are required to provide evidence that the testees are evaluated in a fashion not to be racially or culturally discriminatory. We don't think that this requirement would be aided by a test disclosure provision because the creation of "culture fair" test items is a difficult task to perform in the establishment of standardized tests. Questions which have proven reliability with respect to identifiable linguistic minorities need to be carefuly husbanded. If they were to be revealed, we don't think that they could be effectively substituted. Therefore, the Coalition is concerned that the good intentions of the Weiss bill might be hindered by the actual disclosure provision so that instead of the purported goal of the legislation to ensure equality of educational access, an undesired side-effect would occur: a deleterious weakening of the validity of standardized tests with respect to controls for linguistic and cultural bias. Instead of providing general test disclosure, it is more in the interest of educators concerned with equality of educational access that the test services be required to provide timely, periodic evidence of the non-discriminatory nature of their tests, both with respect to the scaling and scoring procedures and with respect to the actual content of the questions themselves.

The Hispanic Higher Education Coalition wishes to close this testimony with one last observation. The intent of our suggestions is not to put roadblocks in the way of the testing agencies with a view to ultimately destroying the testing industry. Far from it, we see important benefits to be derived from objective tests, including standardized tests. For example, Hispanic educators have been involved with bilingual education for the last 20 years, and longer. Most Hispanic educators readily acknowledge the extremely important contribution that the development of assessment procedures has made to bilingual education. Hispanic educators are usually leary of assessment procedures, and rightly so, because of their potential for bias. And yet, in point of fact, bilingual education as we know it today would not have been possible without the massive—often court ordered, as in the Aspira Consent Decree—development of assessment procedures to determine the language competencies of non-native English-speaking children. Contemporary bilingual education is irrevocably tied to the good services of assessment procedures in determining placement of children within bilingual classrooms. This is all to the good, for bilingual educators often have many sad experiences to recount in witnessing the illegitimate subjective placement of minority youngsters in special education classrooms often for the mentally retarded or learning disabled, simply because of their lack of experience with the English language. Bias among educational personnel is just as real as content bias in tests. Disclosure of test items in the area of bilingual education will serve no useful purpose. Indeed, to cite one example, it is generally acknowledged in the city of New York, that the Language Assessment Battery, the test used to implement the Aspira Consent Decree, is in urgent need of additional, secure versions because a result of the current policy to readminister it over several grades causes children to become "test wise." Because of repeated exposure to the test, children receive a higher paper score than their language competency warrants, and therefore are removed from bilingual education classrooms.

In conclusion, we fully support the movement toward more accountability in testing. We endorse most of the provisions of the Weiss bill. We think that the bill should be modified and strengthened, however, in the particulars that we have described above, in order to better attain its provision for equality of educational access for linguistic minority students. It is because the Hispanic educator probably more than any other minority group professional has been privy to the worst, most biased aspects of the testing process, as well as its potential for accurately assessing competencies and determining placement, that these suggestions are entered in the spirit of a thoroughgoing reform of a service that needs to regain public trust in order to continue to be productive and beneficial to society.

Chairman Perkins. Mr. Weiss.

We will abide by the 5- or 6-minute rule first.

Mr. Weiss. Thank you very much, Mr. Chairman. My voice is such that I am not sure that I can really even carry on for 5 or 6 minutes.

Dr. Keller, I appreciate your testimony. I think that you have really given us some very new and important insights.

You referred in your testimony, and in your written testimony on page 4 you spell out in detail something which I was not aware of and I am not sure if the public at large was aware of. That is, that in fact, not in effect but in fact the graduate record examination itself has a built-in bias in favor of men, that women do not get the same break in that test that men do.

Is that correct?

Dr. Keller. Yes, sir, that is correct. The converse is correct, also. Women do not make the same distribution of scores overall with respect to the quantitative part of the GREs, so that men do significantly poorer. Although they attain a standard distribution, the distribution has a lower threshold you might say.

With respect to the verbal component, women do significantly poorer with respect to the quantitative component of the GRE.

I might add that other standardized tests report out the same results and also have conversion charts for men and for women so that women are their own control group with respect to the percentiles reported out and men are their own control group in analogous fashion.

Another such test is SAT. The same thing at the undergraduate level occurs.

Mr. Weiss. Which is, at the undergraduate level again?

Dr. Keller. Men and women vary with respect to their performance in connection with verbal versus mathematical aptitudes.

Mr. Weiss. But there is no special help provided to women to help them perform as well as men?

Dr. Keller. It is my understanding that such a procedure analogous to the GRE is in effect with the SAT. Scores are reported out with distribution according to gender, male or female.

Mr. Weiss. So that all the arguments that have been made in support of the impartiality and objectivity by the testing companies about the tests and their own attitudes the tests test the same for everybody no matter who they are and where they come from in fact is a lot of nonsense on the basis of this kind of conversion of their own.

I used stronger language than you did.

Dr. Keller. The evidence is there that the scores are reported out by gender differences. That is an uncontestable matter of fact.

So the graduate schools such as my own will look at the distribution of scores for men and analyze those separately with respect to the distribution of scores for women. We don't combine apples and pears and we don't combine men and women with respect to those scores. We analyze them and evaluate them for admission separately.

Mr. **Weiss**. Is there a similar kind of differentiation by way of test record conversions on the basis of ethnicity?

Dr. **Keller**. There is not with respect to the GRE, SAT or any other standardized tests that Educational Testing Service or other kin agencies provide.

There is at least one, however, in existence which is reported in the written version of our testimony produced by Jane Mercer and a large commercial testing house, Harcourt Brace Jovanavitch, that is the SEA, part of which test scores that take into account the ethnicity and cultural background of the testees. That is a sort of IQ test, the SAFPA.

Mr. **Weiss**. The thing that has disturbed me through all these hearings and all the inquiries about it is that when it comes to our concern for cultural bias built in on the basis of ethnicity or race or geography or whatever, the argument is always made that you cannot mess with the—never mind the test itself—you cannot disclose what goes into the process because in some way those tests are sacrosanct and you will be intruding in the educational process. That was Dr. Howe's testimony as I understood it.

Here we have an example that in fact those tests themselves are skewed to get a result the testmakers want which is to give men more of a break in the GRE examples than they would deserve on the examination.

Ms. Rice shakes her head, but isn't that right?

Dr. **Keller**. Yes, a special attention has to be given to the significant statistical difference in the distribution of scores achieved by men as compared to women on the same test with the same test items.

Mr. **Weiss**. Right. I really have only one further question or comment. You know what we have had here all along and the part of the educational testing service and its allies and supporters is an attack on a straw man.

If in fact the bill did what you propose in your first modification which is in fact to mandate the kind of differentiation of norms on the basis of different groups, then they could come in and attack us on the basis that we are trying to tell them what kind of tests to compose and how to score them and so on.

The legislation, and you criticize it, but the legislation very, very clearly prevents and does not touch on that at all. What we simply do is to ask for disclosure and openness.

I don't understand how Dr. Howe or anyone else can suggest that this is an intrusion. We are not trying to tell them what to put in the test. It is not akin to Adam Clayton Powell's questions about what should be in textbooks.

I think a lot of people are being misused and misled by the testing industry who have a very real financial stake in this situation.

Thank you, Mr. Chairman.

Chairman **Perkins**. Mr. Goodling?

Mr. **Goodling**. First of all, I might say my colleague recently made a statement, something about the people opposed to this bill or those people who have some interest in the testing programs or some connection with the testing programs.

Let me tell you, as an administrator and as an educator, I was very reluctant to administer standardized tests because of the misuse that teachers would sometimes make of them as far as determining what a youngster could or could not do.

I have real concerns simply because I think it is an intrusion into something we had better be very careful about before we do it.

My first question would be to Dr. Howe. You have been in this business a long time and very close to education. Do you see an emergency that requires us to move rapidly into some kind of truth-in-testing program from the Federal level before we have an opportunity to see what New York is going to come up with when they begin their program, I think in January, or the National Science Foundation, before they come up with their results of the study they are doing which I believe is due in the middle of 1980.

Do you see that kind of emergency?

Dr. Howe. Mr. Goodling, I would like to take a moment to answer that, if I may, rather than just give a yes or no answer because I think the question is more complicated than yes or no.

First of all, I have a lot of agreement with many points made by my colleague. When he says that testing needs to regain public trust, I think he is right. There is clearly a great deal of concern among educators, among the public in general, on this subject.

When he points to such examples as he has about the misuse of tests that are supposed to be predictive for diagnostic purposes and that kind of thing, what little bit I know about testing—and he is much more an expert on the explicit subject of testing than I am, I am by no means an expert on it—my experience would lead me to think that he is right, that there are abuses of this kind.

My major point, however, is that I do think it is perilous ground for legislation, that to get into national legislation in this field with the hope of correcting those kinds of problems simply by laying things out in the open in the way of test scores and reporting test questions, an openness type of legislation does not seem to me to address those problems.

That is why there is suggested in my testimony a point I did not make before in my summary of it, that you consider appointing a national commission on this subject that would be a distinguished group of interested and informed people with an adequate staff really to look at these questions in the kind of depth that no matter how hard you try in the course of congressional hearings to get interpretations from many sources it is very difficult for you to achieve an indepth understanding.

Therefore, to answer your question, I don't think there is what I would call a national emergency, but I think there is a real problem in public confidence on this matter.

Since I agree with the testimony that has been given that public confidence needs to be reestablished and the kind of specific abuses that have been charged need to be looked at in detail, they require a kind of sophistication and expertise which I am ready to admit that I don't have.

Therefore, again, I welcome this kind of inquiry to give a chance for exactly that sort of ventilation. But I really worry about two things:

One, the adequacy of this sort of legislation to lead to the result my colleague would desire and, secondly, to come back to some things that Mr. Weiss said, I do think that there is implied in this kind of legislation, if not stated directly in it, the notion that the way students are admitted to college in the United States is the Congress business. I think it is primarily the college's business and not the Congress. I think it ought to be kept that way.

We have recently had a major debate here in the Congress on the Federal role in legislation and a very close vote on whether we have a national department of education. This issue of the appropriate Federal role was pretty well reviewed at that time and there was a good deal of controversy on that subject.

When the Congress is saying in legislation that it thinks that multiple criteria is the way to admit students into college, it is entering into the college's business. That is what this says.

It is not requiring that multiple criteria be used, but it is saying that it wants to encourage the use of multiple criteria.

I fully agree that multiple criteria is a fine idea and I am for it. But I am not for the Congress saying that is where it stands. I know of institutions which use open admission systems. Other institutions use other admissions systems and give varying weights to multiple criteria.

I think opening this door of a national expression of wanting to bring about certain results in the business of institutions, many of which are private and those of which are public are controlled by the States and not by the national government, is not a good idea.

Mr. GOODLING. Thank you.

Dr. Keller——

Chairman PERKINS. Let me interrupt the gentleman.

Would you want to recess and go over and vote and come right back?

Mr. GOODLING. I'd like to ask two questions.

What two changes do you suggest?

Dr. KELLER. We suggested that the data be gathered, the data which the H.R. 4949 requests be gathered not only by income level but by linguistic and cultural minority groups, and furthermore that the testing services be required to report scores and norm scores separately from major discernible and linguistic and cultural groups. That is one suggestion.

The second is that those provisions for the disclosure of test questions and answers be deleted from section 5 of H.R. 4949.

Mr. GOODLING. My second question: I am not sure of your age. Are you old enough to know how people were admitted into colleges before we had the kind of national test results now used as part of that admissions procedure?

For instance, if you wanted to go to an elite school and you were a minority or you were poor, what chance did you have?

Dr. KELLER. Congressman, I don't have experiential knowledge of that, but I have the kind of knowledge that accrues to one who is acquainted with the literature.

Let me hasten to say that I think the Hispanic Higher Education Coalition and myself personally recognize many of the benefits of the testing agencies. We in no way wish to curtail or destroy the testing services.

Our suggestions are made in the spirit of reform. We do recognize that to some degree the testing services have provided for access to students who very well might not have been admitted to postsecondary education.

Mr. GOODLING. Thank you.

Mr. FORD. I am sorry I was not here for your testimony, Dr. Howe, but I am drawn to your concern about Federal intrusion into education. This might surprise some people because since we worked together in 1965. I have been intruding constantly from the Federal level in education—that is as an advocate of Federal intrusion.

But, as the chairman of the Higher Education Committee, I have become more and more sensitive to the pressures on us all the time to move into areas that would completely change the relationship of Federal aid to education and Federal concern for equity and access in education. They are constantly trying to get us involved in some type of administrative entanglement.

I would ask Dr. Keller, not necessarily in your capacity speaking for the organization representing Hispanics, but as someone familiar with what happens at Eastern Michigan University and one I am very familiar with, who in Washington would we look to to provide the kind of public confidence which is sought in the enforcement of legislation of this kind?

Laying aside the differences which are involved, if we assume the adoption of some form of legislation, would it be predictable that institutions like yours would be willing to trust a Federal agency to assess the value of the testing they were using for admissions?

Dr. KELLER. I think that many institutions would do so enthusiastically. Many others would certainly accommodate themselves quickly to the idea.

Mr. FORD. Let me interrupt you to say for the record that Eastern Michigan has an outstanding record of recruiting minorities.

Dr. KELLER. Thank you very much, sir.

Mr. FORD. That is evidenced by the dramatic change that has taken place in recent years in the entire makeup of that institution, the attitudes on the campus, and the makeup of the student body.

But that is one of the schools which I would consider to have progressed very rapidly in the last decade and a half, and not much before that. I helped to change its name, as a matter of fact, from Yipsilanti Normal to Eastern Michigan State University in 1961. It certainly was not the change of name which changed the attitude.

But, would we have Federal enforcement which went down to the norm of the recruiting system at our small private colleges in Michigan, or to an Eastern Michigan norm?

Is it likely that we could get some sort of an overall standard which would not go down to the lowest common denominator, and therefore leave Eastern Michigan imperiled in terms of what it does as an institution, while staying conscious of its responsibility to recruit, retain and guarantee the success of minorities? This is a concern that I have.

When you start to talk about setting national standards, it can become very entangled. If I can go to a completely different subject, we used to have the best clean meat law in the country in Michigan regarding comminuted meat, which is meat ground up into sausage or hamburger, until the Federal Government came along to help the American public and passed a consumer bill which set standards for what you can put into sausage. From that day on, we in Michigan began eating snouts and ears in sausage and we never were subjected to that in the past.

In making a national standard, you make it impossible for someone who wants to exceed that standard to do so. When you note the tremendous diversity we have in higher education, you should recall that prior to World War II, I could not have obtained—even if I had the money—entrance to schools which I eventually attended. The system would not let a factory worker's kid get into those schools.

I grew up understanding that doctors' children had a chance to be doctors and lawyers' children had a chance to be lawyers. It had nothing to do with race or sex or ethnicity. It had to do with a very closed system.

Everything we have been doing since the GI bill has been directed toward opening up the system. We have institutions, such as the one with which you and I are familiar, which have taken years to respond to this situation.

I am concerned about what I am hearing in the testimony here about putting the Federal Government in a position where we may, in fact, impede the people who are exceeding without doing very much for the people who are not.

Is that a concern that you view as legitimate?

Dr. KELLER. Yes, Congressman, an extremely legitimate one.

If I may take a minute or two to address myself to your question—by the way, I think that your idea to change the name to Eastern Michigan University was an inspired one, one which has benefited the university greatly.

I believe all colleges and universities in the United States, both public and private, are profoundly disquieted by the manifest relationships that have been reported out, for example, by the admissions testing program or the College Board that show a direct correlation between income level and tests on standardized tests such as the SAT.

Mr. FORD. Does that surprise you?

Dr. KELLER. No, it does not.

Mr. FORD. What is the significance of that? I saw the California study and I was amazed that anybody would seriously say that they had discovered something when they found that people from families with over $30,000 incomes did better in school than people from families with less than a $30,000 income. They do better at everything. That is what Jimmy Carter meant when he said, no, life isn't fair.

I come from a family in which no one finished high school. But, I know what the California study concludes from my work on this committee. Why is it surprising, when you recognize that only 20 percent of the entire population is in that income category above $30,000—probably something less than 20 percent as a matter of fact.

The chances of having parents who have a greater appreciation for the value of education and provide support for educational endeavors in all the ways that are important, is that much greater at that income level.

But I don't understand the significance of relating success on tests, or success in the Boy Scouts or success in any other endeavor, to that income if you don't deny the fact that virtually all the legislation we deal with here recognizes that people from low-income backgrounds start off and continue throughout life with a disadvantage that has to be made up for in some fashion.

Virtually all the programs that we have legislated since the Elementary and Secondary Act in 1965 have been predicated on the assumption that there is a correlation between low-income family background and a need for additional help in achievement of some sort of normal educational experience in this country.

Then, suddenly after all these years of operating on that assumption, not only for the purpose of designing programs but also for distributing most of the money, you take an entirely different approach to that assumption. But now people say we are surprised that people who come from the higher income portions of the country score better on tests and that more of them go to professional schools.

That is not because of the tests. It is because of the good luck they had picking their parents.

Dr. **KELLER**. The tests provide a very special case. As Congressman Goodling, I think, brought up, one of the important arguments and to a certain extent valid arguments of the testing services is that they have provided open access to higher education, irrespective of income level and even perhaps ethnicity.

But to the extent that those tests are not able to discover the potential and the aptitude of students who might be handicapped simply because of cultural or linguistic or social class biases, to that extent it is necessary for the tests to be forthcoming with respect to their deficiencies or inadequacies and it is necessary for those educators who are concerned with the problem to reform those test services and those specific tests in order to make them better.

That is why I am not surprised by the correlation between income level and scores on the SAT. I am not surprised by the apparent—because I don't know the actual details, I don't think the testing services have been forthcoming with them, but there seems to be an apparent relationship between ethnic origin and scores on the SAT, when you read between the lines of table 10 and table 11 of California College Bound Seniors, 1979.

Again, I am not surprised by those relationships but I am deeply and profoundly saddened by them. I think that the testing agencies could do better to provide a quality of educational access, and the suggestions that we make we believe are practical, sound, affirmative action procedures, fair ones, easy to put into practice.

They would help those testing services correct for content bias, bias in the content by modifying the norming procedures. It is easy to do. Why shouldn't we do it? Why shouldn't we help the testing services make their argument more substantial that indeed they are helping to control for the relative states of advantage that can be perceived in different income levels or ethnic origins?

Mr. Ford. But you see, I don't think anyone in this room disagrees with the points that I think you are making. And, I think that Dr. Howe has agreed with you up to that point.

The question is whether you are going to accomplish that by trying to establish some sort of a federally mandated test of what is or what is not the right way to achieve that in a system as complex as we have in this country.

We visited China this year with this committee and discovered that they are doing a very interesting thing. They were very forthcoming in telling us that they abandoned testing in favor of political committees to decide who was going to attend their colleges and universities.

Quite frankly they now acknowledge that they have lost 10 years of development at that level and they are desperately trying to reconstruct their system. They insisted that the first consideration in the future for entrance to their totally public system—a governmental system—will be a student's ability as measured by tests.

On very close questioning, however, we discovered that even there, they are not willing to be that objective. There is going to be an additional subjective test that will be applied at the recommendation of the student's high school or the equivalent of the high school with regard to deportment.

Now I suppose in a Communist country that means whether you are a party member or not, but in any event they are not willing to go the whole route.

I don't think there are many institutions which are basing admission entirely on testing. But I am pleased that testing has a greater influence on the overall makeup of the student body today than it did 25 years ago in this country. Because then the admissions system for colleges was an incestuous, very narrowly based, elitist system.

My problem is: How do we preserve the progress that has been made, and improve upon that progress without having the Federal Government intrude in a way, as Dr. Howe has warned about, which will hold back the people who are blazing the trail and proving that you can do things better.

It is the schools like Eastern Michigan and others around the country that have proven that you can develop a good student body, carry out all the missions of that school, and still meet the ambitions that many people have for improving the lot of minorities and low-income people and still maintain a school with relatively high standards. These are the ones that should be copied.

I would not even like to see a State average in our State, and I don't think you would either. Given the competition for students now, I suspect that if you are warm and have a tuition check in your hand, that there is a college some place in Michigan which will take you no matter how you score on the SAT.

Now, that wasn't true just a few years ago, but today it is virtually that bad. Every time I go near a high school, there are recruiting posters. They are advertising all over the place. Five years ago the shoe was on the other foot.

During this period of tremendous pressure on institutions like Eastern Michigan, the only way they could say no to people was to become very arbitrary with testing. I think that is where a lot of the pressure developed which is now being reflected in the people who are concerned with this legislation.

I share the concerns of Mr. Weiss and all the others. We have just reviewed a very, very arbitrary system which had 15 people for every seat that was available in a freshman class. You have to find some way to ration those seats.

For example, many institutions, particularly the larger ones in Michigan, accepted you or turned you down without ever talking to anyone. You never became more than a punch card. The joke was that if the dean happened to step on your card on the way to play golf, you might get in by accident.

But the truth is that kids were being turned off in the sixties because there was a totally impersonal system adopted. It was in response to the pressure of too many consumers, and not enough available product. That has completely changed. To now develop responses predicated on the extraordinarily wrong things that occurred during that period of time raises some questions with me.

For example, I don't know to what extent there are people who would not like to eliminate dependence on tests because the student enrollment is dropping in some of our small private schools in Michigan. There has to be some standard some place along the line. I don't know the difference between how important a test result is to get into Eastern Michigan University or the University of Michigan, as opposed to Alma College or Hope College or some other small school.

I suspect everybody from Alma or Hope is interviewed individually, and a good deal of subjective determinations go into admissions. There are practically none at the University of Michigan. It is much too big, except in the graduate school.

I don't really understand, after all the years I have tried to wrestle with this, what the relative weight of the SAT is at the entry level if you take away the problems of economic access and access by numbers.

Maybe we are wrestling with problems that are solving themselves as we are here wrestling with them.

Dr. HOWE. Mr. Ford, can I comment on that?

Mr. FORD. Yes.

Dr. HOWE. I think you raised some very important problems. I think it is important to recognize that the type of progress that has been made in Michigan about admitting more minorities to higher education is largely the result of affirmative action efforts.

But different institutions conduct affirmative action in different ways. That affirmative action in admissions is not required by Federal law.

I think we have made some more progress by the congressional insistence that there be equality of educational opportunity in the United States and by the kind of provision of funds to back up that purpose that you mentioned earlier.

But for Congress to try to start writing the exact arrangements how that is going to happen is another problem.

Second, we face a shifting situation in our country's population. By the year 2000, the Spanish-speaking population of the United States, perhaps well before that, is going to be the largest minority segment by a long shot. It is growing more rapidly than any other.

We are going to have to adjust testing and a variety of other things in our schools, and colleges and institutions are going to have to make major changes to be fair to a growing population with a language and indeed a culture that demands some institutional adjustments.

So, again, I say that calling those kinds of matters to the attention of educators is very important, getting them on the agenda for consideration and making sure they stay there is very important. But I do believe that efforts to try to particularize how institutions should respond on this from the National Government is full of perils.

To repeat what I said before, I think it opens the door in legislation which does not require institutions to do anything, as Mr. Weiss quite properly says. It opens the door through which all kinds of other changes in education are then likely to be coming before the Congress, before the Congress for more detailed legislation, and to me that is unwise.

You said, Mr. Ford, that you had regarded me in earlier years as an advocate of Federal intrusion into education and that now you thought I had changed my stripes.

Dr. Howe. I call your attention to the fact that I testified vigorously against Mr. Powell's effort to get into the curriculum. I will do so again if it comes up.

I think that that kind of issue, while not directly raised here, is indirectly raised here, and that is why I included that matter in my testimony.

Chairman Perkins. Mr. Buchanan, any questions?

Mr. Buchanan. Thank you; no, Mr. Chairman.

I will read very carefully that testimony which I missed.

Chairman Perkins. Mr. Peyser, any questions?

Mr. Peyser. No; Mr. Chairman.

I also arrived late, and will reserve any questions until a later time.

Mr. Weiss. Dr. Howe, I don't know how you stand on the creation of the new Department of Education. I opposed it very strongly exactly for the reasons that you have expressed about concern for Federal intrusion into education. I take second place to no one on that.

Dr. Howe. I waffled on it.

Mr. Weiss. I make no comment on that.

Chairman Perkins. All right.

Thank you, gentlemen, very much.

Dr. Keller. May I just address myself very briefly to the last questions that Congressman Ford brought up?

Chairman Perkins. Go ahead, but let's make it brief. We have a lot of witnesses here.

Dr. Keller. I think that the question that you are asking is why should the Federal Government intervene in this situation.

My colleague, Dr. Howe, has agreed with me that there is a climate, perhaps a critical one, of public mistrust with respect to the testing services, and the litany of complaints is exceedingly long. Given that climate in the spirit of a thorough going reform of the services, not only does Congress and the Federal Government have the right to intervene but it would be wise to do so in order to shore up those important services that the testing agencies give to us.

In addition, you observed with respect to my suggestions that identifiable linguistic and cultural groups be normed differently. You made the following observation, that you agreed with it, I believe, and you thought everyone else in this room probably agreed with it.

I am gratified by that comment, and I sincerely hope that is true.

However, I am not sure, and my reading of the situation at this point is that the testing services make the argument that the scores can be reported out as if there were only one distribution of scores irrespective of income, sex, linguistic and cultural parameters, even though in point of fact, as Congressman Weiss has mentioned, they actually separate the distribution with respect to sex.

Our contention is they should follow that precedent and separate out the scores and norm separately those students of identifiable linguistic minority backgrounds.

Mr. FORD. I have to disagree with you only to that extent. I agree with your general proposition, and continue to agree with the general proposition, but when you get to the specifics, you get to the problem of dealing with national standards.

When you are trying to predict the likelihood of success in law school where the entire program is in English, and where the basic testing system in most of the law schools is by essay questions which depend very, very strongly on your ability to communicate well in the English language, I cannot conceive of why you should test somebody in a language other than English. This is questionable especially when it's clear that it is not general intelligence here that you are testing for.

It is the probability of being able to respond to that course of study no matter how arbitrary it may seem. I use Puerto Rico, for example, where we discovered that children who go through their entire educational experience, then reach health professions education and must switch to all-English study materials.

The explanation given to me was that there aren't Spanish textbooks or teaching materials, nor are there trained personnel to teach in Spanish. So they have to make a total transformation by language. Thus, in Puerto Rico, it would make sense to give the test in either English or Spanish.

It would not make sense, it seems to me, to test for admission to a professional school in another language where the language was going to be in English.

It's apparent to me that if you have a system, you don't do anything to change the system, you do something about how somebody penetrates the system.

You have two standards about how you get into a professional school, and two standards about how you progress in the professional school, and then you ultimately meet a blind test called the bar examination.

One group is not going to do as well as the other. It has nothing to do with their intelligence But I predict to you that if we start reporting test results, SAT, LSAT's and the rest on the basis of race and ethnicity all over this country, we will have kooky professors all over this country saying, "I told you so; these people are inherently inferior."

If you want to push civil rights a generation back, you start publishing Government-accepted national figures showing that certain groups score lower than others. For every intellectual who will grasp that as proof of the fact we have to improve the opportunities for those people, there will be 10 people listening and say that is proof of the fact that these people are inferior.

It sort of surprises me that some of my friends over the years who typically have been concerned with civil rights don't see that as a threat. They apparently are not listening to the winds that are blowing across this country. Hate is just under the surface, but it's never been so willing to bubble out and slap back at the people we have been trying to help in this country as it is now; and I see in that kind of approach a threat that perhaps people in Washington don't hear.

Maybe Michigan is a little different than others.

Dr. KELLER. We entirely agree with you with respect to the language in which the tests should be written for mainland United States. It should be English. We are not saying that there should be separate tests created in Spanish for the mainland United States.

The example that we use of Puerto Rico is with respect to necessary research that has to be done to determine how much of the tests are really measuring language and culture rather than the purported aptitude such as law.

However, it was shown by Professor Hoffman that the tests are created in a written and artifical language with misleading inappropriate ambiguities, and we suspect that fluent English speaking Hispanic Americans might be prejudiced more than other populations within the United States by such artificial language.

That is one of the reasons why we ask for separate norming. There are plenty of people at this time, Congressman Ford, who claim that women are inferior or blacks are inferior or Hispanics are inferior. At the same time the precedent has already been established, women and men are reported out from those tests differently; different percentiles are reported out.

We don't think that any more kooky professors than those who already exist have or will jump on that fact with respect to women; and we don't think that they will jump on that fact with respect to identifiable linguistic and cultural minorities.

But one thing that could happen is a net increase of students entirely capable to enter college and professional schools but who have been turned back simply because to some extent the tests have been measuring the linguistic parameters and that distortion could easily be looked at by separating the linguistic minority population from the mainstream population in norming separately.

We think it would be a good way to implement the equal access philosophy of U.S. education and the equal opportunity clause of the 1964 Civil Rights Act.

Mr. **Ford**. We are still not in disagreement. How would you like to be trying to get at the root of the problem of someone whose basic language is not English and who is not doing as well at your institution, and then be confronted with somebody with a list of statistics prepared by a Federal bureaucrat that proves that the group that you are dealing with can't score anyhow, who, for whatever other reason, does not want to see you make progress in that area.

That is what the problem is. It's not the problem with the inadequacy of the institution but the inadequacy of a category of people. They are going to have a Federal stamp of approval on the phoney figures they throw around. They can't prove the assertion they are making, and that is the only defense you got against those people.

You said a little while ago that you had confidence that the very fact the Federal Government would certify these tests to be fair would instill confidence in the public.

I am afraid that that is not an advantage. We don't know whether this Federal agency would make the test fair, but they would appear to be fair because the Federal Government said they were.

At that point, if you were stuck with that test, you would really be stuck, because then you get Uncle Sam putting on his "Good Housekeeping" seal, which has been put on bad information, because I don't trust the Federal Government to be able to do any better than we are doing on an ad hoc, State-by-State or institution-by-institution basis now.

If they do a lousy job, and if they do what you suggest is true, people believe that the Federal Government certifies that tests are fair, they will then have undue confidence in the fairness of those tests; and you will never make any progress beyond that point.

Somebody over here who doesn't know what he or she is doing says something is good, and because the Federal Government has tested it, they believe it is OK. We have this problem in accreditation now. We have a big argument going on whether the Office of Education ought to be accrediting accrediting agencies.

The Office of Education says the fact is that they accredit anybody who applies to become an accrediting agency. The result that produces is that people are led to believe that because the Office of Education said this is an accredited accrediting agency that, therefore, it in fact, meets the requirements that people would expect, and then display that on their literature and advertise for students, and the students go to that school thinking in some way the U.S. Office of Education knows whether or not that is a good school.

Mr. Buchanan and I have been wrestling with that for months and have all kinds of political pressures coming at us, but we can't get anybody to suggest what the solution is.

We wouldn't have a problem if we didn't have people believing that the Federal Government was doing something to protect them that it does not do.

We have so many examples of well-meaning Federal attempts to protect the consumer that in fact have a reverse effect. People believe that ever since we have had the Federal Deposit Insurance Corporation that they can trust the banker to take care of their money, and they never have to worry.

That has never been true, but the purpose of that legislation has been fulfilled by rebuilding in this country after the Great Depression confidence in the banks but that confidence is now misplaced by an overconfidence because the ordinary citizen doesn't take the time to find out what the Federal Government is really guaranteeing you against, and many people are far beyond the scope of any help from a Federal Government when they are in a bank that may have 50 percent of its folio in New York bonds and they are in Michigan.

The people didn't run on the bank because they believed that Uncle Sam would protect them; and too many times we hold out from Washington a suggestion that we are doing something that will protect you and so people don't watch out for themselves. They don't listen to groups who try to tell them, look, you can't trust Uncle Sam to look out for your interests, and you have to trust us, because we are more interested.

This is where I have the concern. I am not really concerned about the kind of amendment you are talking about and the specifics of what is or is not the real problem at this point. I just can't understand how the Federal Government is going to improve it.

Dr. KELLER. I don't think that the Federal Government could possibly certify the fairness of tests, but it can, and I think should, take strong affirmative actions to get the test services to be more forthcoming in disclosing information periodically on a timely basis, showing what and how they are addressing themselves to the problems of content but, as I say, that reflects negatively on linguistic minorities and certainly I think it is in order for the Federal Government to require a different norming process for such minorities.

That will not certify that fairness of the test, and the Government would never want to do that, because tests can always be improved. Measurement will always be imperfect to some degree and improvement will always be in order. However, you can make an extremely timely contribution to American education at this time which is suffering from a climate of public mistrust.

Congressman Ford, I wish to thank you very much for your kind remarks with respect to the activities over the last 15 years that we have been undertaking at Eastern Michigan University.

Dr. HOWE. Could I just say, Mr. Ford, that I share your dilemma about the business of accrediting institutions. I remember when I was Commissioner, I came up against the schools of mortuary and embalming, and I didn't know how to advise the Congress on the subject of how these institutions should be properly handled. I concluded it wasn't a good field for the Commissioner to be in, or the Congress, and that is the way I feel about this subject here.

I would have many worries about the testing services themselves going into reporting their data strictly on ethnic grounds or ethnic classifications. But for the Congress to say that that is what they should do and should norm their tests in those terms, it just seems to me extremely unwise.

I come back to my suggestion about an in-depth national look in an effort to build confidence and develop information on this subject. This kind of suggestion that my colleague has made could well be examined in that context and examined without the weight of national authority behind it. The national and international implications of this Government legislating that the testing information about youngsters in school shall be collected on an ethnic basis just seems to me worth thinking about a long time before you put anything like that into legislation.

Mr. BUCHANAN. Will the gentleman yield?

Dr. Howe, I would appreciate your elucidating a little, if you would, on your proposal for a commission. We had a witness yesterday who recommended that, in lieu of legislating in this field, we use our leverage, with perhaps—with the club of legislation in the background. However, it does seem to me that getting at the truth is what the author of the Weiss bill has, at least partially, in mind.

I wonder if you would elucidate a little.

Dr. HOWE. It does seem to me, as a generalization, when we face very complex and difficult national issues, that we frequently resort to this device and other countries do.

The Royal Commissions in the United Kingdom are for this kind of purpose. They are frequently used in the field. They are frequently used in the field of education. We have not used them in exactly the same way here in education, but we have had some significant commissions, and we have got one working now on international studies appointed by the President and I think it will provide some useful suggestions, some of which may result in legislation.

But it seems to me that when we have a highly technical area loaded with frustrations for various institutions, on the one hand, and for various groups on the other, that the legislative process is probably wise to allow itself to be insulated from action by that kind of commission.

I do think that many of the serious difficulties that many people see in this area can receive longer term, more thoughtful, more analytical attention, and can be ventilated over a period of time. It may be that such a group would recommend legislation. I would doubt it, however.

I think that as a basic proposition, really indepth legislation in this field does change the balance of the Federal, State, and local roles and institutional roles in education, and that is why, without having any negative comment to make at all about the evident issues of unfairness that have been called to your attention that there be by other witnesses, I do worry about rushing into this field.

Mr. BUCHANAN. Thank you.

Chairman PERKINS. Let me compliment the panel. You have been most helpful to the committee and we appreciate your appearance here this morning.

Dr. Howe, we will see you back here again. We appreciate your being back here and we will see all of you again. Good luck to you.

Now we will ask the panel to come around headed by Mr. Albert Shanker, president, American Federation of Teachers, who has been before the committee many times; Dr. George Jackson, National Association of Black Psychologists; Dr. D. Kay Clawson, dean, College of Medicine, University of Kentucky; Mr. David Gastfriend, chairman, medical education committee, American Medical Students Association; and Dr. Lloyd Bond, research associate, learning research and development center, University of Pittsburgh.

Also, Mr. Kenneth Drexeler, president of the student government, North Senior High School, Great Neck, N.Y.; Dr. Clark R. Cahow, associate professor, history and assistant provost, Duke University, consortium on financing higher education; and Mr. William Geer, Jr., principal, Frances W. Parker School, Chicago, Ill.

Our first witness this morning is Mr. Shanker. Let me welcome you here again, and you have been with the committee many times.

I have got to be out for just a few moments in the back room, but you may start now. You have many other members here.

Excuse me, we have got a colleague from your State that wants to introduce you, and that is Mr. Peyser, who is one of our most outstanding members.

Mr. Peyser, you go ahead.

Mr. PEYSER. Thank you very much.

As we welcome the entire panel, I particularly want to welcome you, Mr. Shanker for the years that you have been working with this committee and the committees on education. I think the testimony on this issue—an issue that is really developing more and more awareness—is pointing out what this legislation is involved with and what is happening in New York State.

We will be most anxious to hear what you have to say. If you would start at this point, we will be delighted to hear you first and then question you.

I will be glad to yield to the gentleman from Michigan.

Mr. FORD. I ask for this time to apologize to the panel. I have an 11 a.m. commitment in line with my responsibilities as chairman of the Subcommittee on Postsecondary Education, and I am leaving someone here to catch the highlights which are set forth. I hope that I will be able to direct some questions to you later in the hearing.

Mr. PEYSER. Thank you.

Mr. Shanker?

[Prepared statement of Albert Shanker follows:]

PREPARED STATEMENT OF ALBERT SHANKER, PRESIDENT, AMERICAN FEDERATION OF TEACHERS, AFL-CIO

I am Albert Shanker, President of the American Federation of Teachers, AFL-CIO. I am here today representing over 520,000 teachers, paraprofessionals and health care workers who have a profound interest in the future role of both aptitude and achievement tests in our society. Our members have an interest in the proper use of quality tests not only because of their effect on the futures of children they teach, but also because of their use in partially regulating entry into the teaching and health professions.

I believe that there are two simple questions that must be asked and answered in considering the merits of the two testing bills you have before you, H.R. 3564 and

H.R. 4949. The first is, what are the pressures and problems they attempt to address? The second is, do they provide an appropriate response?

These are not simple questions. Certainly I do not have all the answers. Proponents of this type of legislation don't either. Only Monday the New York Times reported that New York state senator Kenneth P. LaValle, who marshalled similar legislation through the New York State Legislature, was considering recommending changes in it because of unforeseen problems that had arisen since that bill became law in July 1979. I now believe that the haste with which the legislation was passed in New York was a mistake, even though the AFT affiliate there supported it. I urge this committee to carefully consider all the implications of the legislation before you, since any mistakes that are made will affect the extremely delicate business of measuring educational achievement throughout the nation.

Let me point out that I find some of the consumer-interest concerns reflected by some supporters of this legislation very understandable. Much of the press reporting on education would have us overemphasize the use of tests in making key decisions without fully considering their limits. The American Federation of Teachers has stated its opposition to this practice, as have some members of this committee. But our Executive Council will shortly reconsider the specifics of positions we have taken in the light of what has already happened in New York State and with a clearer view of the possible consequences of similar legislation.

The basic purpose behind the development of objective aptitude and achievement tests was to provide us with an objective standard of quality, that could be used in conjunction with various subjective measures like student grades and personal interviews, in making decisions about admissions to various postsecondary undergraduate and graduate programs. An objective test like the SAT, for example, can just as well be used to "discover" the bright student who has done poorly by subjective grade standards as to raise questions about the possibly inflated subjective reports on a grade-A student. There is every indication that the public continues to want objective and comparative test standards. The continuing support for elementary and secondary minimum competency tests at the state and local level are clear evidence of this. The push for this bill must be balanced by taking these sentiments into account.

Today the issue of tests is being considered in two, highly polarized ways. On the one hand is the continuing push to weigh test scores more and more heavily even further in decisions related to grade promotion, high school graduation and postsecondary school admissions? On the other is the attack being waged against tests by those who cannot bear to face the implications of what their results tell us. Neither extreme position serves us well. Some testing advocates, failing to consider the real limits of test use, are too willing to rely on test information alone in making educational decisions. Test detractors on the other hand, readily dismiss the value of all comparative test information, and along with this the importance of what tests expose about unequal educational opportunity. I would hope that in considering these bills, the members of this committee will take care that the legislation they recommend does not have the effect of siding with either extreme. My fear is that it will seriously impede our ability to test and to use test information well.

Let me discuss these pressures in relation to the effects of what I believe to be the most controversial provision of either bill, Section 5 of H.R. 4949, which requires test publishers to provide test takers with the test and their corrected answer sheets. Advocates of such a measure assume two things—first, that faulty test items will be exposed and open to criticism by students and others; and second, that the student will benefit from seeing where he was wrong. I believe that the first of these intentions is what moves supporters of this bill the most and that its potential effect must be examined very carefully.

Constant standards

What will the effects of this exposure be on tests, when the obvious purpose is to subject each item to rigorous scrutiny and at the same time make it difficult to use that item again? While it must be acknowledged that tests, and the individual items in them are sometimes flawed, I believe that the potential negative consequences of politicizing test scrutiny, accompanied by the remaking of tests, far outweigh whatever benefits may derive from these processes. Let me explain why.

The most basic benefit from good educational tests, consistently applied over long periods of time, is that they provide us with the only constant standards we have over time for assessing educational progress and achievement. Constant standards are essential because without them, we cannot be certain whether children are learning more or less than they did in previous years and we cannot be certain whether some methods of helping children learn are more effective than others. These are vital social facts: educators, legislators, and other policy makers must have them to do their jobs properly; parents, children and taxpayers in general also need them and have a right to them as citizens and as consumers.

That right is threatened by the provision for the blanket disclosure of test questions. Good intentions notwithstanding, such provisions are dangerous because they severely undercut the key method used to maintain constant standards. This method is known as equating and it involves including samples of previously used test questions on each new version of a test to make sure that each new edition of the test is no easier or no harder than the old ones. Different tests require different methods of equating and the restrictive provisions of these laws may make constant equating difficult if not impossible for some tests. This is a very serious matter. Any insistence on the disclosure of all or more test questions after each test administration necessarily and inevitably jeopardizes test constancy and the essential educational truths it permits us to see.

We are aware that some test-producers have promised to try to maintain test constancy in spite of these blanket disclosure requirements, but we find these vague promises totally unsatisfactory. We would also question their ability to produce an endless series of quality tests with all new questions in every single edition. We doubt that the testing agencies can or will maintain test constancy given the factual circumstances confronting them.

Quality tests

There is another consideration that must be raised here, the issue of quality. Once the process of test development is speeded up, and at the same time each test is subject to the rigorous scrutiny of test critics, I believe that the quality level of the tests will also be subject to dangerous fluctuation. Some will argue that particular items are inappropriate for a given region of the country, or a particular sex group, or a particular ethnic group, or for people who live in rural areas, ad infinitum. I believe that particularist groups will argue for items peculiar to their own experience or orientation and this will have the effect of erroding a common quality standard.

Too much Federal control

This legislation's potential effect on test quality and constant standards are the most compelling arguments against it, but there are others well worth noting. Why should the requirements it outlines be a matter for the federal government? Do we really want the federal government to determine that "no educational or occupational admissions test which tests knowledge or achievement (rather than aptitude) shall be graded (for purposes of determining the score required to pass the test for admission) on the basis of the relative distribution of scores of other tests subjects." (H.R. 3564)? Do we really want the federal government defining admissions standards in such detail? Besides, what is wrong with evaluating comparative performance, and selecting the most competitive candidates?

I have predicted that the creation of a new and separate bureaucracy for education at the federal level will bring increased federal intrusion into educational matters that are more appropriately decided by individual states, school districts and universities. I cannot help noticing that many supporters of this intrusive legislation are also among those who advocated that separate Department of Education, and who are waging a campaign against testing.

I opposed the creation of a national test when it was proposed by Admiral Hyman Rickover two years ago, and I oppose federal government regulation of tests today for many of the same reasons. The reports required of test publishers and of the Commissioner of Education by this legislation actually amount to a series of federally mandated conclusions about the effects of the tests given. These requirements are virtually designed to produce conclusions that will amount to national verdicts on tests, and I view this as one step short of granting the federal government the authority to approve some tests and not others—in effect, to nationally control tests.

Why, for example, should each testing agency provide a comparison of the average score and percentiles of test subjects by major income groups only, unless the legislation's authors think this is the only meaningful set of relationships for test publishers to look at?

Why should the Commissioner report to Congress on the relationship between the scores of test-takers and income, race, sex, ethnic and handicapped status only, unless it is assumed that these are the only meaningful sets of relationships he and the Congress both should look at?

These kinds of provisions in the law amount to the legislation of a particular research methodology. Why not look at test scores as related to family size, for example, or class size, or school size, or family composition, or number of books in the home, or any of the many other variables researchers suspect may relate to test performance? Should federal legislation insist that a federal official view limited data in a limited way? Obviously the answer is, no. These types of provisions are extremely dangerous. They are precisely why we should be very cautious about legislating in this area.

Cost, availability, and other arguments

There are other arguments which must be considered, many of which have already been presented to this committee:

The cost of test development, and ultimately to students will inevitably go up. Should the application of disclosure requirements be made to employment testing and oral examinations, as is provided for in H.R. 3564, the cost to government and to private industry would magnify as well.

The number of times a test is offered will inevitably go down as a result of cost factors.

While some in New York—ourselves included—had a tendency to dismiss these arguments as simply the contrived opposition of the testing industry, their validity must be given some weight when we now see that test publishers are voting with their feet by deciding not to administer some tests in the state and not to forward test results to New York Colleges and Universities.

The federal government does not require what this legislation would require for its own civil service examinations and armed services examinations. In fact, its testing records are specifically protected from disclosure. Since this legislation has clear implications for existing government policies their apparent inconsistencies should be thoroughly examined.

While the intentions of this legislation's authors are good, I do not believe that the ramifications of the enactment of either of these bills has been adequately thought through. I welcome the discussion that has surrounded them as a great contribution to education of the public and educators on the subject, including that of my own organization. But, after reviewing criticisms that have been made of this legislation by others, and after witnessing the initial confusion that is accompanying the enactment of similar legislation in New York, I am convinced that federal legislation of this type would be dangerously premature. My concern for the preservation of test quality and constant standards, and my fears regarding a drastically altered federal role in education are behind my recommendations to you to delay quick consideration of these measures until all of their potential hazards have been thoroughly reviewed.

STATEMENT OF ALBERT SHANKER, PRESIDENT, AMERICAN FEDERATION OF TEACHERS, AFL-CIO; DR. GEORGE JACKSON, NATIONAL ASSOCIATION OF BLACK PSYCHOLOGISTS, ACCOMPANIED BY ALVIS ADAIR; DR. D. KAY CLAWSON, DEAN, COLLEGE OF MEDICINE, UNIVERSITY OF KENTUCKY; DAVID GASTFRIEND, CHAIRMAN, MEDICAL EDUCATION COMMITTEE, AMERICAN MEDICAL STUDENTS ASSOCIATION; DR. LLOYD BOND, RESEARCH ASSOCIATE, LEARNING RESEARCH & DEVELOPMENT CENTER, UNIVERSITY OF PITTSBURGH; KENNETH DREXELER, PRESIDENT, STUDENT GOVERNMENT, NORTH SENIOR HIGH SCHOOL, GREAT NECK, N.Y.; DR. CLARK R. CAHOW, ASSOCIATE PROFESSOR, HISTORY AND ASSISTANT PROVOST, DUKE UNIVERSITY, CONSORTIUM ON FINANCING HIGHER EDUCATION; WILLIAM GEER, JR., PRINCIPAL, FRANCES W. PARKER SCHOOL, CHICAGO, ILL.

STATEMENT OF ALBERT SHANKER, PRESIDENT, AMERICAN FEDERATION OF TEACHERS, AFL-CIO

Mr. SHANKER. Thank you very much for this opportunity to be here and testify on this important matter.

I am Albert Shanker. I represent the American Federation of Teachers, which has over 500,000 members in elementary, secondary, and higher education.

We are very much concerned with the issues which have been raised around this proposed legislation. In hearing some of the testimony this morning, I am reminded of something which came up a few years ago when Sidney Marland was U.S. Commissioner of Education.

I remember at that time the Office of Education proposed to invest some money in a school district in Brooklyn which would develop a new set of tests which would be normed in such a way that the students in that particular district which chose the same range of distribution as students did nationally.

This was a district which had great educational problems. The students were not achieving as well as the norms. It was a district with heavy minority concentrations.

What I found was very interesting in that situation was when the Commissioner proposed that such a test be created, there was an outcry from the parents of children in that community who came down to Washington and said, "We don't want a test that is going to tell us that our children are doing as well as other children when we know that they are not."

Finally that project was abandoned. I would be happy to go back into the archives of the Washington Post and the New York Times to pull out those clippings and to see how rapidly a government ventured into the testing business for a few weeks, how quickly it got out. It was driven out, I am sure, by the very people that some of the people in the Office of Education thought they were going to appeal to and help.

Before we get into a few of the specifics here, I have submitted written testimony, which I am not going to read. I will just make a few points.

I believe that someone is misreading the current American temper. I do not think that the No. 1, or 2, or 3, or 5 items on the minds of the American people is a whole list of complaints against tests.

I think most people out there think that they took tests, and most of them never thought that the tests were not fair, or they were biased. When they did well on the tests, they knew the tests were good. When they flunked them, they thought that they hadn't prepared themselves.

Most people still think that way, that it is a very, very small group of people in this country who have developed this sophistication about what it is the tests measure, don't measure, and what the difficulties are.

The big public concern in the country today is not with the problems of testing. It is with the declining standards, and everyplace you go, whether it is local or national or at the State level, there is a demand for a greater use of objective tests.

There are problems with that, we know, but I think that that context is very important, that we are considering here developing a cure for a disease which most people don't recognize.

As a matter of fact, the cure for the disease that we are working on is likely to create problems in terms of achieving what the public wants to achieve, which is a tightening of standards and a raising of standards.

If you create confusion in administration of tests and everything else at a very time when you are trying to improve standards and tightening up, you start raising questions about this test and how much additional work has to be done and what it means, and so forth.

You begin a long process of changing the measuring instrument and you will have a public out there that is going to believe that the reason we are changing the measuring instrument at the very time when scores are going down is because we don't want them to know what the results are.

I do not believe that we have a public in this country that is going to spend billions of dollars for education, elementary, secondary, and higher education, and deprive itself of information as to whether that money is buying anything, and whether we are getting better or worse from time to time in terms of the delivery of that service.

So I hope that the points I raise are taken in that context of what I believe is a concern.

I happen to agree with that concern. I think it is important before we go into all the negatives of testing that there are some positive things, positives that may be destroyed. If others get into this active business of bringing about some major modifications—and let me start with one very clear and simple one.

The existence of tests pretty much the same tests which are given over a long period of time creates one of the few reliable pieces of knowledge we have in our society about how well our schools and our students are doing.

The interesting thing when the SAT scores declined, you did not have a bunch of people saying they didn't really decline. They are measuring something.

The tests have changed. Every expert in the country who looked at those said yes, there is no way you can deny it. There are things that students were able to do *x* years ago that they are not able to do now. I wish the scores had gone up.

I wish we didn't have the bad news, but we have very little reliable information in the field of education. As all of you know, you propose almost anything and you will get a whole bunch of people here interpreting one thing or another.

It is very good to have a piece of legislation that no one has contested; namely, that whatever those tests measure students are not doing as well as they once did. It is important to have that.

I think secondly those tests are used as one piece of information and evidence as to whether or not students should be admitted into certain colleges or into programs or in terms of one piece of legislation, perhaps in terms of employment relationships, et cetera.

I don't know of anyone who has come forward and said the only evidence that has been used is an objective test. It is one piece of evidence, and I will submit that my experience as a teacher, not in higher education, but in elementary and secondary, that for every time a test was used to classify a student, and perhaps deny him some opportunity, I would submit that my experience is that it worked 20 to 1 the other way, that if you had a student who did not test well but who responded well in class and who did well on the work and everything well, the tendency of the teacher or the institution was to question the validity of the test because you had the evidence of the student's performance right in front of you.

On the other hand, if you had a student who was not doing particularly well but the test showed that that student supposedly had great ability, you started working on that student and sit down with the student and say, "This test shows you are really very able and you are not doing as well as you should."

You talk to the parents and we have to give them help and provide outreach, guidance, so forth.

Now, I think that what we are caught in here is a conflict between what I think is the overwhelming pressure of the American public now to use tests more and to tighten up against a coalition of forces.

Some of it is part of a very good consumer movement but you have to go very carefully and make sure you are not destroying old values with the new ones you are trying to create.

But also there is a movement in this country that is kind of naive that believes if you get rid of the tests you automatically create quality.

If you get rid of the yardstick, everything is equal, and I am not making that up. I have been in major national debates here in the Capitol sponsored by the National Institute of Education and others where prominent organizational leaders stood up and said tests rank people, so that somebody is first and somebody is second and somebody is third. That makes people unhappy, et cetera.

I would say that we ought to be rather careful that we are not overresponding to the views of, I think, very small groups that believe that achievement should not be a purpose in the schools, that the schools are designed for happiness or for play or for pleasure or for all sorts of other things.

I certainly think we should try to make those in our institutions happy and try to make learning as pleasant as possible. But we ought to keep our eye on what schools are about and what education is about. Achievement, of course, is a main part of it.

Now, it is our belief that a program which would require the kind of disclosure which is being proposed here will destroy the knowledge base that now exists. If you keep creating new tests each time because you can't give the one that was given last time because you had to release it, you are not going to be able to tell the public 10 years from now whether the SAT's went up or down because it will be a different test each time.

We think that is a very big price to pay, and we ought to think long and hard before paying that price.

We are impressed with the evidence that making up new tests constantly will substantially increase the costs.

We are impressed with the evidence that as these tests are opened up and made public, that we will definitely further the industry of cram courses and that we will in a sense be determining the curriculum of many of our schools by making the passage of these tests the major subject of study for quite a period of time.

I would say that it is certainly open to question whether any of these things are going to help the rich or the poor. I dont think the evidence is in, and anyone who is going to support this on the basis that they know what the answer is, I think they are developing answers which are not there yet.

Now, just a point on this question of the way these tests are supposed to be reported in terms of race, in terms of ethnicity, in terms of sex, income, and so forth.

If you are going to have a research program, I would agree with Dr. Howe, who just spoke about the dangers of that, and associate myself with the remarks of Congressman Ford.

If you want to conduct research into what it is that makes students not do as well, you ought to find out how often these students move from one school to another, while they are with the elementary school and secondary school, what the psychological and social services were in the community, you might throw in something about the medical care they received and the question of how many hours a week they watch television.

If we are going to do some of the research, let's do some research. But I am very worried that in deciding that these are going to be reported on an ethnic basis, for example, that we are really determining something.

We are really saying that if year after year given ethnic groups do poorly on certain questions, therefore those questions must be culturally biased and must be thrown out. That was one of the debates I had with Dr. Marland when he was Commissioner of Education.

What we are going to do is get the companies and testing outfits to keep throwing out those questions which tend to separate. Some do better, others do worse, et cetera.

I think it is kind of a nonsensical approach. It may very well be the questions of simple arithmetic or questions in the use of the language in a sophisticated way, that these skills because of past discrimination and recent entry into the country, language difficulties and so forth are not equally distributed at the present time, but that does not mean that they are not valid questions to ask.

If we go through this process of getting rid of these questions, what we are really doing is destroying standards and an adequate testing program and really politicizing it. We will end up, I think, the public is not going to lose faith in testing because of this.

They will lose faith in government and lose faith in education because they are not going to believe that we are maintaining the standards that they want.

I want to finally point to what is happening in New York State, that testing publishers have made a decision there not to give certain tests, not to provide results within that State.

There are already members of the legislature in the State who are rethinking legislation and there may very well be some changes there. I will stop at this point just saying that I am not ready to say that there are never under any circumstances, that we should absolutely stay away from any legislation.

I think it might very well be that after lengthy study on this that there is some greater scrutiny to which these tests can be subjected, but I don't like the rush. The issues are very complex and the consequences are pretty serious, if we make a wrong mistake.

I would hate to sit here with a list of actions which have been taken in recent years by the Congress in the name of advocacy and consumerism which boomeranged with all the unanticipated consequences that are sitting there.

There are as many unanticipated consequences as in any of these other fields, so I am here to urge that you couldn't look at this and move very slowly.

Mr. PEYSER. Thank you, Mr. Shanker, for your testimony.

The way we are going to proceed is to go through the panel and then the member will have an opportunity of questioning.

The next witness is Dr. George Jackson, National Association of Black Psychologists.

Dr. Jackson?

STATEMENT OF DR. GEORGE JACKSON, NATIONAL ASSOCIATION OF BLACK PSYCHOLOGISTS, ACCOMPANIED BY ALVIS ADAIR

Dr. JACKSON. My name is Dr. George Jackson, National Association of Black Psychologists, 1417 16th Street, NW., Washington, D.C. I speak for the association. I am a professor at Howard University.

At the school of social work where I am currently a professor—excuse me, Dr. Adair is also with me from the association and will be among the persons who will answer questions, if there are any, for us.

At the school of social work we turn out approximately 125 students a year with MSW's. We refuse to use standardized testing as a means of looking at our students. Our students we would stack up against any group of students throughout the country.

One of the reasons, many of the reasons why we are successful in what we do is because we do not countenance falsehood.

It is our contention that to suggest that the standardized tests are able to measure the aptitude of an individual, the aptitude of an individual, even as a partial predictor, even as a partial predictor, is a fallacy.

We believe that it is time for the Federal Government to move into the arena because the test companies have been recalcitrant. They have not behaved as honest scholars. Indeed, they have attempted to place within the marketplace insidious comparisons based on false assumptions.

The Association of Black Psychologists in 1969 decided to call for a moratorium on the testing of oppressed people in this country. It was not a response to those individuals who said that black people were inferior.

What we were addressing is the inherent nature of the tests itself. It was our contention that—and it is our contention now—psychometry in its current condition cannot and does not measure what it purports to measure.

If you are talking about prediction, which is what SAT, GRE say they are talking about, measure an aptitude which they are then able to predict from, if you are talking about prediction, it is incumbent upon the person who has a theoretical formulation for doing this to demonstrate not only prediction but a logical relationship between the predictor and the criteria.

You can use any variable as a predictor—skin color, hair color, union membership, et cetera—in tems of how people behave. So it is incumbent to talk about not only prediction but also to talk in terms of the logical relationships.

If we talk in terms of prediction, and we use multiple equation, as you enter more predictors into the equation each of your components decreases in its predictability.

If you are talking about using correlations for cause and effect or relationships because of a statistic called the coefficient of alienation which uses the standard deviation and square root of your variables.

You need to have over a 0.90 correlation before you can talk about any significant degree of variance which is describable. It being describable to your predictors under considration, or as we put forth in 1969, there is an issue of what we call the moderating variable which suggests that when you make correlations between, say, test scores and GPA in racist institutions, such as are all over this country, that you may be talking about not only a correlation which is high because the two variables are related but because there is a third variable, a moderating variable, a racist curriculum.

These and other things are what led us in 1969 to call for a moratorium for testing the disadvantaged. I will not go through the whole list of considerations.

I would remind you again that we also suggest that the assumption of the norm curve as a measurable, as a measure of continuous variables of ability across ethnic groups, we regard as a fallacy.

So at the very least it seems to me if companies are compelled to place their questions on the table annually, it will foster. It will compel a new technology within the psychometry industry, and possibly we might then through being able to get feedback and knowing what the instrument is, maybe eventually they will stop developing what might possibly be called the achievement test, which they are not either, instead of aptitude tests.

Some years ago, many years ago, in fact, the New Jersey Commission for the Blind made it essential for a blind person to take the SAT and to score at some level, I don't know what, in order to get a scholarship to pay for blind persons' tuition, and so on.

I don't know how it happened, but somehow I was 2 years in the university and suddenly someone noticed that I hadn't taken the SAT, and they called me up and said you have to come in and take the SAT in order to get a continuation of your scholarship.

I was getting all A's at that time, excepting in religion, and went down to take this test, which I later learned I flunked, so there I was moving into my junior year in college, full scholarship, A average, and they were telling me the only thing we can let you do is take the test over again.

I asked, "What am I going to do?" I was told, "Well, we don't know." I said, "What can I study?" They said, "We have a Braille booklet we will send you which gives you some sample items, but we can't tell you what to do," so I did nothing.

The next year I took the test again in Braille and this time I got a letter saying, you passed it by such a margin, it can only be called fantastic. I don't know how I passed it or flunked it, and I was unable to prepare myself.

There was another blind student who went through high school with me who used to do mathematical problems in his head. Unfortunately, he took the test before he tried to go to college. He flunked it.

He went to a factory to work. I am not trying to suggest that people shouldn't work in factories, but he was interested in going to college.

You might suggest that this is a one-person example. It is a case of one. My experiences in higher education as a black psychologist, as a musician, and as a politician at times indicates that this is not.

In Newark, N.J., for example, in 1958 on first trial, not one black student got into a college that year; that is, to the multiple criteria.

We talk in terms of the test being an objective indicator. We would submit that the test is a psuedoscientific replacement for the objective or the subjective racist attitudes which we black people confront, and we would submit that the tests have been used to exclude us from admission.

Yes, it is understood that there are ethnic differences. On most of your psychometric tests, the mean performance for blacks as compared with whites is one standard deviation below the mean, below the the white group in most of your standardized testing.

We would submit to you, we understand that we do not have education at the level that we would like to have it. We understand that, but in this notion that the test brings the message and, therefore, don't kill the messenger.

We submit that is a fallacy. We submit that the messenger is trying to kill us, and we are not going to let it kill us.

I think that we have to understand that it is not subjectivity which determines what test items will be pulled, nor is it subjectivity which finally determines what a cutoff point will be.

It has never been determined. It has never been determined by a college or university that they could show a relationship between SAT, their grade points, and success in the world of work or performance.

It has been suggested that SAT's and other standardized tests reflect what is going on in the schools. We say it does not. It does not mirror what is going on in the schools. In other words, the tests do not measure what they purport to measure.

Therefore, to use them is a fallacy and places a travesty and dishonesty upon the young people of our country. In Howard University we have a newly developed doctoral program 5 years in existence. We do not use the GRE.

We have students now, only one student has been lost for academic reasons. Standards at Howard University continually go up, and so does the level of our students' performance.

It continues to go up. When I was the dean of academic affairs of Essex County College in Newark, I was able to move, with the help of the faculty, our nursing program from almost losing its accreditation because of the scores that our graduates were getting on the nursing test.

In 1 year we moved that group to receiving the highest score in the State, and there was no relationship between the tests which they recommended we use as a predictor and the outcomes.

What we did was we taught the students, and we taught them not only the content, but we taught them that they are important, something which the tests say they are not.

At Seton Hall University, where I was the director of black studies, we brought in the students who had the so-called nominally low SAT scores. Those students should not have been able to perform in college if the tests were measuring an aptitude which we say the tests did not.

Within 2 years you could not distinguish those special students from other students in the mainstream of the university. And this kind of activity where you have faculty and you have people believing in the fact that they can undo perhaps some of the damage that was done in the public school system makes it possible for people to function.

In conclusion, if we had used as a criterion the black people who graduated from our black colleges and universities who were taught by black scholars who had a heritage of coming from an educational system which in fact was on a par with anything that we had seen in this country, if a criterion group of that type had been selected, perhaps maybe we wouldn't have called for a moratorium. But the test companies still do not do that.

I would have to agree or disagree with Mr. Shanker, and I am glad to disagree with him because I can think of no person who tried harder to hurt black people through *Bakke*.

It is very interesting that he and I are on opposite sides of this issue, but we cannot judge issues by the company people keep. It is just that I never thought I would have an opportunity to say that so close to him after seeing him on television for many years.

Finally, in 1975 the Association of Black Psychologists, recognizing that the climate would not be conducive to continuing the moratorium, called for more than a moratorium. We called in various articles and publications—we called for truth in testing laws.

Through the offices of the National Association for the Advancement of Colored People we did sit many times at the table with ETS and College Boards. They did acknowledge that these tests behind closed doors are culturally biased.

We called for more than a moratorium. We heartily endorse all aspects of this legislation.

Chairman **Perkins**. Let me thank you very much.

Our next witness is Dr. D. Kay Clawson, dean, College of Medicine, University of Kentucky.

STATEMENT OF DR. D. KAY CLAWSON, DEAN, COLLEGE OF MEDICINE, UNIVERSITY OF KENTUCKY

Dr. **Clawson**. I appreciate the opportunity of testifying here.

The College of Medicine at the University of Kentucky was established in 1960, specifically to address the issue of the doctor shortage in the eastern part of the State and in some rural counties that had been traditionally underserved.

The purpose of my testimony today is to try and show you how the standardized MCAT has been an advantage to us in the admissions process, and to caution you about what I think would be the dilatorious effects of this particular law which you go through on the admissions process to not only our college of medicine but to a number of other ones in the country.

Let me say something about the admissions process into medical education. Admissions policies are established by faculties in the various colleges.

This committee usually composed of basic and clinical scientists, practicing physicians from the community, students, alumni, and knowledgeable lay individuals are the ones that make the admissions decisions to our schools.

Chairman PERKINS. Have the students in your institution always had input in the setting up of standards?

Dr. CLAWSON. We have always had students on our admissions committee, yes, Mr. Chairman, that is correct. We believe that if you look at the varying criterion used, the MCAT is only one. We give consideration to the accumulative grade point average.

We also use the MCAT in a positive way because we believe it maximizes the objectivity and insures consistency and makes the process more fair to all individuals.

Let me tell you a little more about this process because we are very well aware that a B grade in one college may have the same standard as an A or a C in another, and we have no way of knowing this except in the prejudicial mind of people who may be sitting on this committee or a faculty as to how they might weight a college.

In addition, we also know in a college there are those instructors who are called hard graders and also those who are called easy graders and we have no knowledge about this.

The MCAT as it is now constructed and as it now works has been an advantage to us in that we have been able to take students that might not for past reasons be considered on the objective sense of their grade point, their past performance to be competitive, to undergo the medical curriculum, and on the basis of scores that they have received on the MCAT have made it so.

I would illustrate two from the recent class: One, a boy from a mountainous Pike County in Kentucky who had scored very poorly, less than a C grade, in college. He went into the military and came back to college. He got better than a C. He started working as an orderly and then a respiratory therapist in a hospital environment and had an excellent track record and still wanted to go to medical school.

I can say without his performance on the MCAT I don't think an admissions committee would have looked at his academic performance and passed in a good position.

Another man in Russell County had a B-minus level performance but in grade point levels for admission he would not have been competitive and yet the MCAT which showed him scoring in the top 20 percentile was clearly the thing that tipped the scale and allowed him to the admissions process.

So I have numerous illustrations like that I could go to, but I do feel this particular test has been a great advantage to us.

Let me say something about the MCAT itself. We have been talking about a lot of other tests that I know about, but I don't know the details. I know American medicine was very concerned about the old MCAT and through our American medical colleges we have contributed over $1½ million to do research in and try to make a better tool, even though an imperfect tool, as a standardized test, and that is what I call our new MCAT which was administered for the first time in 1977.

We think it is a much better test. Why we cannot open up the pool of questions is very simple. We have decided in the process of doing this that we want the future physician to have a very broad education. Therefore, we have said that the questions we will test on will be from the normal first year of college categories and have selected from that area the questions and material that we think is broad based and essential and have the student coming from a small college have the same opportunity as a student from a large college.

We hope we have done this in the process and the pool of questions is very, very small. If we open it up, then we would have to expand the pool and we would go back to a situation where students would be spending years studying certain scientific disciplines in order to pass the test and try and cover the whole horizon rather than getting a broad education.

We think that is a wrong way to go. For that reason we are very pleased with what we have now been able to accomplish in this particular regard.

What would happen if the test was opened up, I can tell you probably what would happen in medicine. I go back to the days when it was quite different, when standardized testing was not as readily available, and when you got in by either who you might know or because your father was from a particular background or a particular school.

At this point in time, we do have a good tool, one of many. The grade point average and the MCAT combined in our institution only make up 60 percent. We look at humanitarian communities, we look at how the student has performed in social, religious and other groups. We look at their background in athletics, art, music, and other things. We give extra points for those coming from underserved areas and underrepresented groups in medicine.

So we think we have a fair process of which the MCAT as it is now constructed and as it is now applied is a very valuable tool in doing what I believe the proposer, Mr. Weiss, wants to accomplish and that is to getting underrepresented and underprivileged ones into our medical school.

If we are to lose it and go back to locally administered tests, they won't be as good. This will be a burden and a problem for students. If we rely on the subjective judgment of the committee members sitting around the table, I have been through that for many, many years and I know how helpful it is for me if I look at a student's record and somebody or somebodies are, I think, unconsciously discriminating against some of our students who for economic reasons or for cultural reasons have elected to go to one of the small, less prestigous colleges.

There is a psychologic thing in academia that might not give them the same level of credibility as going to the more prestigious institutions, and I can say, now wait admissions committee, take a look at this. Look how this student really has proven he can perform coming from one of these institutions.

I think it has been a great tool for us to accomplish the goals that the bill wants to accomplish and for that reason I am very much concerned about the deleterious effects this legislation as it is particularly proposed right now would have on our admissions process.

Thank you.

[The prepared statement of D. Kay Clawson follows:]

PREPARED STATEMENT OF DR. D. KAY CLAWSON, DEAN, COLLEGE OF MEDICINE, UNIVERSITY OF KENTUCKY

Mr. Chairman and Members of the Subcommittee: I appreciate this opportunity to present testimony on H.R. 4949, "The Educational Testing Act of 1979." I am Dr. D. Kay Clawson, Dean of the College of Medicine at the University of Kentucky in Lexington, Kentucky. As you know, Mr. Chairman, our College was founded in 1960, because Kentucky was a doctor shortage state, and one of our primary missions has been to increase the number of physicians practicing in the Commonwealth, particularly in the eastern, mountainous area which has traditionally been underserved.

My testimony will be directed at the deleterious effects the proposed legislation would impose on our medical school's admissions process and will attempt to show how the use of the standardized MCAT test has been of great advantage to us in getting qualified rural and non-traditional students into medical school.

The University of Kentucky College of Medicine requires, as do 99% of this nation's medical schools, the careful consideration of the New MCAT results as an integral part of the review and evaluation of an applicant's credentials. If enacted, the test disclosure provision of H.R. 4949 would greatly impair the confidence placed in the reliability of the New MCAT test results, and might eventually eliminate their use entirely. Many scholars and educators have put in countless hours of extensive research and development to produce this very specialized examination.

A few words about the medical school's admissions process may be in order. Admissions policies are established by the faculties of the various colleges. The selection process is performed by a Committee appointed by the Dean, consisting of basic science and clinical faculty, students, practicing physicians from the community, and frequently, knowledgeable lay individuals and alumni. To be fully aware of the implications this proposed legislation would have on the utilization of the New MCAT, and ultimately on the admissions process, it is important to understand that the evaluation of each applicant's credentials is a comprehensive and exhaustive process. No single criterion determines acceptance to medical school. Since the medical student, and ultimately the practicing physician, must possess a broad range of personal and academic qualities, the admissions decision is based upon the Committee's best judgement of the applicant's possession of these qualities. Additionally, it is imperative that the process insures that these judgements are executed in a manner which: (1) maximizes objectivity; (2) insures consistency in our review and ultimately in our decision, while allowing for the uniqueness of each individual as they establish their total record; (3) makes full and appropriate use of all indicators of the applicant's previous academic achievement; (4) responds to current and future societal needs; and (5) enhances the diversity of the academic and personal backgrounds of the students in our medical schools to provide a heterogeneous output of future physicians capable of providing a broad scope of personalized services, as well as adequate numbers to enter into research and academic pursuits.

At the University of Kentucky, we give maximum weighting to the cumulative grade point average. However, we feel the MCAT is essential also to insure that there are not inequities in the GPA measurement, and to identify students who may, for one reason or another, have had a poor semester or year, but whose total knowledge base is significantly above what their grade point would indicate. I would like to amplify on these two points. First of all, a "B" grade point performance in one College may signal the same level of performance in another College as an "A" or a "C." Furthermore, there is no consistency even within a College; some instructors are "hard graders," and other instructors are "easy graders." The MCAT, as it is now constructed and administered, calls the attention of the Admissions Committee to potentially strong students who, because of grading practices, may be overlooked. Furthermore, and most important for us in Kentucky, many of our students go to smaller, less recognized institutions, and there is the constant danger that subconsciously, or even consciously, Admissions Committee members may weigh far more heavily the academic achievement of students who have gone to the traditionally well-known institutions; hence the very bright and capable student who for economic or social reasons has selected one of our smaller schools is penalized. The weight placed on the MCAT gives this student an equal break with students who have attended more prestigious institutions. Secondly, it is frequently alleged that too much reliance is placed on a single test such as the MCAT. However, I would suggest to the Committee that a student is equally, or at times more, severely penalized by weight placed upon a less well designed evaluation tool in the final examination of a required course, than on the well constructed and administered MCAT. The MCAT allows a student who has had a bad experience to prove that he has a greater knowledge base to prepare him for the study of medicine than is indicated by his grade point average. Two specific cases come to mind which, I believe, will illustrate my previous comments. Both students were applicants to our College of Medicine and both were accepted for admission.

The first student is from mountainous Pike County. This person entered the military after attaining less than a "C" average in one and one-half years of college. After his military service, he still did not perform that well in college, but he began working full-time in jobs as an orderly, pulmonary technician, and respiratory specialist. He completed all college requirements for medical school, but with a low grade point average. His competitive MCAT scores, however, were the added cognitive criteria which increased the Admissions Committee's level of confidence in his academic ability and potential. Other supportive documents, letters of recommendation, and positive interview reports contributed to his selection.

The second illustration I would present to further illustrate my point involves a student from rural Russell County. This student entered college as an engineering major, one of the most difficult courses of study in the undergraduate curriculum. His performance in the first two years was safisfactory—about the middle "C" range, and his final two years were in the B+ range. His cumulative performance, nevertheless, was too low to be competitive with other medical school applicants. Once again, the Medical College Admission Test proved invaluable in assessing his academic credentials. The candidate's MCAT scores placed him in the top 10-20 percentile of all candidates taking the exam. Without this standardized examination, he would have had little chance for admission.

While I have stressed the weighting of the grade point average and MCAT scores for this testimony, I want the Subcommittee to recognize that in the admissions process, many non-cognitive characteristics are identified and given almost equal weighting to the cognitive dimensions. These characteristics include: (1) humanitarian, cultural, service, or religious activities; (2) premedical recommendations; (3) interpersonal, group, club, or social accomplishments; and (4) vocational pursuits, artistic, athletic, or professional achievements, residence in physician shortage areas, and minority status.

Before I amplify more on why I believe this law would destroy the unique value of the New MCAT and severely compromise fairness in the admissions process, I would like to reiterate several points made by Dr. John A. D. Cooper. The medical schools of this country, through the Association of American Medical Colleges, have invested approximately $1.5 million in an unprecedented review of the admissions testing program and the research and development necessary to have a more relevant, cognitive test. This has resulted in the new MCAT. As you know, it was first administered in April of 1977 and was the result of a comprehensive review and rating of the relative content prerequisites of over 150 medical school faculty members. It is extremely important for the members of the Subcommittee to recognize that all subjects covered in a normal one-year introductory course in biology, chemistry and physics are not tested. Only specific topics within each of these science disciplines have been selected and these topics are published in order for students to prepare for the test. It is equally important to emphasize that questions designed to test the topics within the disciplines must be limited to the depth of coverage normally involved in a first-year undergraduate college course. The decision to limit the content coverage was made consciously and deliberately to discourage students from taking all of their undergraduate experience in specific science disciplines, the antithesis of a broad undergraduate education. When specific questions and answers must be disclosed, a severe constraint is placed upon our ability to develop and administer a reliable test. The requirement of H.R. 4949 to disclose test items would, thus, prevent out testing program from maintaining its present direction and specificity. If proposed legislation is passed, it would certainly destroy the important elements and direction the New MCAT test has taken, it would undermine the confidence the Admissions Committee members place in it as a tool for admission, and it would certainly make the admissions process more subjective. Furthermore, this requirement could potentially lead to the development of local tests which would increase the burden on candidates with limited resources and there would be no assurance of the same standard of quality. While the MCAT is only one dimension of the admissions process, it is an important one in assuring fairness in the admissions process to all students. It is well known that such things as personal recommendations presented to the Admissions Committee are sought by the applicant; therefore the process of self-selection most certainly influences the nature of the submitted recommendations. Grading and evaluation practices vary considerably among and within undergraduate colleges and universities, and it is totally impossible to think that Admissions Committees can be aware of the various differences. Moreover, preconceived ideas in the minds of Admissions Committee members could prejudice them against very desirable applicants. When discrepancies arise and reduce the importance of any one of these sources of information, it may present an applicant in an unfavorable light. Without the reliability of the New MCAT, I predict that medical school admission committees will be slowly forced to rely on more subjective and highly ambiguous information in reaching admission decisions. I, therefore, respectfully urge the members of the Subcommittee to carefully examine the claims made by those in support of this type of legislation. There is an abundance of cliches regarding admissions tests that is almost totally unsupported except by anecdotal evidence. I do not see how the provisions of this bill will achieve the desired results of those proposing it. Minorities and those of nontraditional backgrounds will lose the opportunity the MCAT has given them to demonstrate cognitive abilities at a level which enables them to proceed into and through medical school. I strongly recommend that this Subcommittee report negatively on H.R. 4949. I sincerely thank you, as Chairman, for the opportunity of expressing my great concern regarding this piece of legislation.

Most sincerely,

D. Kay Clawson, M.D.,
Dean, College of Medicine,
University of Kentucky.

Chairman **Perkins**. All right. We will get around to the questioning in a few minutes.

Our next witness is Dr. David Gastfriend, Chairman of the Committee on Medical Education, American Medical Student Association.

STATEMENT OF DAVID GASTFRIEND CHAIRMAN, MEDICAL EDUCATION COMMITTEE, AMERICAN MEDICAL STUDENTS ASSOCIATION

Mr. **Gastfriend**. Thank you for that error, Mr. Chairman.

Mr. Chairman and members of the committee, first of all I am grateful for the opportunity to testify on the Educational Testing Act of 1979, H.R. 4949.

My name is David R. Gastfriend. I am a resident of the State of Pennsylvania, a graduate of Haverford College, and presently a medical student in the fourth year at Jefferson Medical College in Philadelphia.

I speak about testing with a relatively fresh recollection of the scholastic aptitude test (SAT) the old medical college admission test (MCAT) and as recently as 2 weeks ago the tests of the national board of medical examiners (NBME).

I also serve the American Medical Student Association (AMSA), as national coordinator of its committee on medical education. This committee convenes the student representatives to curriculum committees at all U.S. medical schools.

AMSA is comprised of over 20,000 dues-paying medical students and is dedicated to the improvement of medical education, health care and health care delivery. It holds the belief that access to basic health care is not a privilege but a right.

Our commitment and expertise in the health care problems of rural America is recognized by HEW in that MASA has been the contractor for the summer program of the National Health Services Corp.

In the past 4 years, AMSA has been increasingly concerned over the issue of standardized testing, particularly with regard to the NBME which publishes the instrument most widely used in medical licensure.

We hold a seat on the board of directors of the NBME and have participated vigorously in recent investigations of these tests. This testimony, however, will be related primarily to tests utilized in medical school admissions, i.e, the MCAT.

Let us begin with the assumption that the role of this legislation is to protect, not the standardized tests, but this society which makes use of the tests. Therefore, consider not whether this bill will have an adverse effect on publishers of the tests, or even on educators.

Instead, we address ourselves to whether the present situation creates adverse influences on education, or whether society and those seeking education stand to benefit from this reform.

While the Association of American Medical Colleges (AAMC) contends that the MCAT is not a sole criterion in the selection of America's future doctors, this is only a partial truth. Since the test scores are utilized to screen applicants out, for many this is the equivalent of a sole criterion.

I personally have known many otherwise highly-qualified premedical students, I could tell at least two stories about graduate students in the biological sciences who were sufficiently qualified to be employed by medical schools to teach medical students, yet who were clearly told by advisors and admissions officers to retake the MCAT because their scores would preclude acceptance.

So I think there are flaws in citing what may turn out to be exceptions, even though they might have personal relevance and we may benefit more by citing comprehensive studies that have been made on the tests.

This is a problem on MCAT because 2 years ago the test was changed. I would like to go into what problems were cited as the need for that change. I think the opinion of students may be borne out in the next 2 or 3 years by the studies that the change is not as significant as the Association of American Medical Colleges might have us believe.

The AAMC holds that the tests are essential to determine who will be the best doctors, that they are designed to predict how well a student will perform in medical school and later as a practicing physician.

Actually, medical educators have disputed this among themselves for the past two decades,[1, 2] and recent research has shown that there is no significant correlation between MCAT scores and performance as a physician.[3]

One study which I cite, conducted at the University of Missouri-Columbia School of Medicine, reported, concerning admissions, and I quote:

> Typically, GPA and MCAT scores are used for this screening process. As the screening device becomes more restrictive, possibilities exist that students with other desirable characteristics will be eliminated from the applicant pool.
>
> Neither of the cognitive measures, college GPA's or MCAT scores, correlated significantly * * * with internship letter ratings after four years of medical school. (Exhibit A.)

Another study from the University of California, Berkely, stated:

> It must be concluded that ratings of internship performance for these students were not related to and could not be predicted by MCAT scores * * *.[4] (Exhibit B.)

This report even found that MCAT scores tended to be negatively related to ratings of the young doctors by their superiors. Therefore, the point is made that emphasis on these tests is not useful, and may even be detrimental in the selection of our future physicians.

There is further evidence on this issue. A major study of the MCAT prior to its revision, conducted by the AAMC itself, found that MCAT performance was significantly related to race, and stated that:

> An admissions policy based exclusively on GPA's and/or MCAT scores would proportionately excude minority applicants from medical education.[5] (Exhibit C.)

This is troubling, particularly since these tests historically emerged in the search for equality in education, and objective standards held the promise of rewarding ability, selecting purely on the basis of merit, without regard to racial or any other prejudice.

Now, we learn that the objective instrument itself may have an inherent bias. Thus, many admissions officers found it necessary to create separate criteria for the admission of those students who could best serve the needs of the society. Meanwhile, those students selected by the test as the best and the brightest may not have been the best for society.

What are the benefits of this proposed reform, however, because since these tests have been shown to have problematic and possibly detrimental influences on the selection of new physicians, we will let that stand as a primary concern, and not elaborate on secondary issues such as whether the bill will injure the reliability, validity or economic expediency of the tests.

However, regarding fiscal implications, I would like to state as a medical student that when a test will determine the course of the

---

[1] Gottheil, E., and Michael, C.M., Predictor Variables Employed in Research on the Selection of Medical Students. J. Med Educ, 32: 131-147, 1957.

[2] Gough, H.S., Hall, W.B., and Harris, R.E., Evaluation of Performance in Medical Training, J Med Educ, 39: 679-692, 1964.

[3] Murden, R., Galloway, G.M., Reid, J.C., and Colwill, J.M., Academic and Personal Predictors of Clinical Success in Medical School, J Med Educ, 53: 711-719, 1978.

[4] Kegel-Flo, P., Predicting Supervisor, Peer and Self Ratings of Intern Performance, J Med Educ, 50: 813-815, 1975.

[5] Waldman, B., Economic and Racial Disadvantages as Reflected In Traditional Medical School Selection Factors, J Med Educ, 52: 961-970, 1977.

rest of my life, the price of it is the least of my worries, and I have yet to hear a fellow student haggle over the cost of an exam in the hopes of a discount.

Indeed, let us consider whether the constituency, the student and the public, stands to gain from this reform. There are three student needs which are addressed by this bill:

One: A need for feeback;

Two: A need for equal opportunity for test preparation; and

Three: A need for input and participation in the evaluation and policy determination process. Let me begin with feedback.

When you think about feedback, you might wonder, for what God forsaken reason would a student seek feedback on a test after sitting through 2 solid days of nearly 1,000 multiple-choice questions? What could be more tiresome than to then review all the answers you might ask.

The problem here is that you are looking on from the perspective of one who has done well. Look again, but from the other perspective: Upon seeing your scores you now realize that you may well fail to gain admittance to the college of your choice, or you may fail to gain admittance to medical school entirely this year.

If you are dedicated to the aspiration of becoming a doctor, you certainly will want that feedback badly. You will want to know the weaknesses in your knowledge base, your pacing, your comprehension and interpretive skills. We know students want this feedback. We know because they pay for it.

The many students who seek out commercial preparative courses, paying as much as $500 to learn this same information, often from pirated or verbatim-memorized test items, in hopes of 50- to 100-point improvements on SAT or MCAT scores, these students and many more seek the very feedback this legislation would provide.

However, at the present time, only certain students can afford these luxuries. Release of this information would tend to equalize the opportunity for students to prepare for these tests since every student would have access to a pool of standard questions and correct answers from which to pretest and review performance.

Finally, there is the problem of accountability and diversity of input into the policy determinations of test publishers. Educational testing service (ETS) reports that "* * * authority concerning its operations is dispersed among sponsors, committees, professional associations, staff and institutional users."[6]

Yet, considering the immense effect of this corporation on the lives and careers of students, students are notably omitted from any such role at ETS. We believe that both students and the public have a necessary and legitimate role in policy determination among educational test publishers.

In AMSA's experience with the NBME, this role is not easily accomplished by the presence of a token representative on a corporation's board. Access to information may provide a more tangible route.

I would venture that with the information that this reform would make available, individual students, student groups such as

---

[6] Fremer, J. and Irby, A.J., Why Should Some Tests Be Secure?, paper presented at the Spring Conference of the National Cosortium on Testing, Arlington, Va., April 1979.

AMSA, independent research teams, consumer interest groups, et cetera, would find it easier to examine the policy manifest in these tests and these groups would be able to make more active and valid contributions in this area.

To summarize the points I am bringing to you, there is evidence that the MCAT is not essential to the selection of good physician candidates, although it may too often be used as if it were essential and its use is not helpful in selecting future physicians, although it may be detrimental.

This is the essence of our concerns. And economics, reliability, and validity only serve to cloud the issue.

Students clearly want feedback on these exams which influence the course of career and life. The Congress should act to equalize currently available advantages in commercial cram courses by providing actual test responses and corrections directly to students.

The public needs more data and better means of scrutinizing the test industry so as to address the very types of discrepancies discussed above. For these reasons, the American Medical Student Association supports the principles of H.R. 4949.

Thank you.

[Material submitted for the record follows:]

EXHIBIT A

Academic and Personal Predictors Of Clinical Success in Medical School

Robert Murden, M.D., Greg M. Galloway, M.D., John C. Reid, Ph.D., and Jack M. Colwill, M.D.

Abstract—Admissions interview data and college academic credentials of five classes of University of Missouri-Columbia medical students were evaluated as possible predictors of clinical success. Those students who were judged by admissions interviewers to have high levels of maturity, nonacademic achievement, motivation, or rapport were approximately two to three times as likely to receive outstanding internship recommendations as those without such personal characteristics. Undergraduate grade-point average had a smaller but nevertheless significant relationship with clinical success as measured by internship letters. These data suggest that additional emphasis during selection upon applicants' personal characteristics would have enhanced the clinical success of these students.

Annually admissions committees at each of the nation's medical schools face the difficult task of selecting a class from a huge pool of highly qualified applicants. In most schools this process begins by screening out candidates who have lower grade-point averages (GPA) and Medical College Admission Test (MCAT) scores. Committee members then interview the remaining candidates. Finally, admissions committees select those candidates perceived to be most outstanding, using as selection criteria data on both academic and personal characteristics obtained from application forms, letters of reference, and interviews.

The effectiveness of selection processes should be demonstrated through high correlations of each of the selection criteria with success in medical school. Many studies, however, have demonstrated little or no correlation of various admissions parameters with academic success in medical school.

Typically, investigators have reported low correlations of GPA with performance in the first two years of medical school and even lower correlations with clinical performance. Despite these low correlations, GPA in most studies has been the best single predictor of subsequent medical school performance (1, 2). [redacted] (3, 4).

Dr. Murden and Dr. Galloway, formerly medical students at the University of Missouri-Columbia, are currently resident physicians at the University of Texas Medical Branch, Galveston, and the University of Iowa College of Medicine, Iowa City, respectively. Dr. Reid is professor of education and Dr. Colwill is chairman of the Department of Family and Community Medicine at the University of Missouri-Columbia School of Medicine.

The greatest value of the MCAT test has been to screen out those with low scores who are more likely to have academic difficulty (5). However, with today's surplus of all but minority candidates, the issue is how to select outstanding candidates rather than how to screen out high-risk applicants.

Relatively few researchers have examined the relationships of admissions interview data and medical school performance. Gottheil and Michael (3) described an unpublished study by A. W. Hurd, who reported a correlation of .39 between interviewer's perceptions of applicant's personality and first-year medical school grades. Gough, Hall, and Harris (4), studying 81 students, found a low correlation of .11 between overall admissions interview ratings and overall grades in medical school. Richards and Taylor (6), using 65 Utah medical students, reported correlations from .16 to .32 between admissions interview ratings and grades in medical school. Johnson (7) obtained a correlation of .17 between overall interview ratings and cumulative medical school grade average of 399 Syracuse students. Mensh (8) reported that neither interview data nor college grades correlated with clinical grades at UCLA School of Medicine.

At least to a degree these low correlations may result from methodological difficulties. For example, admissions committees seek candidates with high GPA. If all selected applicants had a GPA of 4.0, the correlation for that group with any outcome measure would be zero. Thus, when selection is based upon a predictor, the selection restricts the range of the predictor. Consequently, later efforts to demonstrate a correlation of that predictor with success can only yield lowered correlations.

Further, multiple factors probably contribute to medical school performance. If leadership were such a factor, an admissions committee might select a student with strong leadership characteristics even though his GPA might be low. Success of this student in medical school would further contribute to low correlations of GPA with academic success. Another contributor to low correlations is low reliability of either the predictor or the criterion measure. Such might occur when GPAs at different colleges reflect different levels of achievement (9) or when the performance criterion is unreliable.

Thus, the best assessment of the importance of various predictors would require random selection without regard to the predictors to be tested, high reliability in measurement of the predictors, and high reliability and validity of the performance measure. It is little wonder that correlations in past studies have been low.

While the absence of a significant correlation of an admission variable with medical school performance may not rule out its value as a selection variable, it does suggest that selection of students with still higher scores on the variable will not increase further the quality of the class. Conversely, the presence of a significant correlation, even when small, suggests that added emphasis on the predictor may well increase the quality of the class.

The purpose of this paper is to report the relationships of measures of personal characteristics and of college academic performance with clinical success, measured by internship letters, of University of Missouri-Columbia medical students four years later.

Emphasis is given to the degree to which the specific personal characteristics of "maturity," "nonacademic achievement," "rapport," and "motivation for medicine," as judged by admissions interviewers, were related to clinical performance. Previous studies of the importance of the admissions interview have used overall interview ratings rather than ratings of specific personal characteristics. Significant correlations between ratings of specific personal traits and clinical performance four years later would support the value of the interviewing process and justify efforts to develop more objective measures of these predictors.

Also of interest was whether measures of personal traits could assist admission committee members in selecting the strongest candidates for interview. Typically GPA and MCAT scores have been used to screen candidates for the interview. However, some authors have suggested that raising the GPA level necessary for acceptance, without consideration of personal traits, will not necessarily increase the quality of the class (7, 10). A screening process based both upon cognitive factors and personal traits might enhance the quality of students brought to interview.

Procedure

Admissions data, including college academic performance and interview ratings, were analyzed for five medical classes (458 students) entering the University of Missouri-Columbia School of Medicine from 1966 to 1970. Sixteen students were dropped for academic reasons, withdrew voluntarily, or had not graduated by June 1974.

During this period all applicants from Missouri and a screened portion of out-of-state applicants were interviewed individually by three members of the admissions committee. Each interviewer used the students' total file, including application materials, college transcripts, MCAT scores, and letters of reference.

Each candidate was rated on the traits of maturity, nonacademic achievement, rapport, and motivation for medicine. The interviewer rated each of these characteristics on a five-point scale from −2 to +2 based upon the degree to which he felt the characteristics either enhanced or detracted from the candidacy of the applicant. Those admitted fell on a three-point scale from 0 to +2 with 90 percent being rated either 0 or +1. Students rated zero on any of the four traits were regarded as having "nonenhancing" qualities with respect to that trait, those rated from 0 to 1 were regarded as having "enhancing" qualities, and those rated from 1 to 2 were viewed as having "highly enhancing" qualities.

Personal characteristics were not precisely defined for the committee. However, weekly three-hour meetings of the committee to discuss and select applicants created informal operational definitions.

Nonacademic achievement was broadly defined for the committee as "leadership activities, accomplishments, and other measures of nonacademic success." Success in college extra-curricular activities was usually the basis for committee members feeling that nonacademic achievement enhanced the candidacy of the applicant.

In this paper the performance criterion is the internship evaluation letter prepared on behalf of each graduate by a 10-member Internship Advisory Committee. This committee estimated a student's future success as an intern by classifying the student into one of four categories: "A" recommended with enthusiasm, "B" strongly recommended, "C" recommended with confidence, and "D" recommended with reservation. Since only seven of the 442 were in the D category, they were added to the C category in this study.

The committee met weekly during fall months to review and modify draft evaluation letters prepared by individual committee members. Each committee member advised eight to 10 students and had responsibility both to counsel those students regarding internships and to prepare the initial draft of the evaluation letter. The committee as a whole, having available the draft letter, the student's file, and usually a personal knowledge of the student, voted on the level of recommendation. Inasmuch as the committee used clerkship evaluations as part of the information to rate students, these clerkship evaluations correlated between .68 and .77 with internship letters. The students were rated on their clerkships in the following areas: clinical knowledge, basic understanding of medical principles, manual and personal skills, attitudes, and desirability as a house officer.

Further, correlations of admissions data with each of the clinical clerkship measures were nearly identical to correlations of admissions data with the internship letters. Thus, the findings reported here would have been similar had clerkship measures been used as performance criteria instead of internship letters. Correlations in this paper are Spearman rhos.

Findings

Table 1 lists in decreasing order correlations of personal traits and academic credentials with clinical performance. As generally noted in previous studies, overall correlations are small. [redacted]. On the other hand, ratings by admissions interviewers of the personal characteristics of maturity, nonacademic achievement, motivation for medicine, and rapport correlated significantly with the internship letter rating ($p = .0003$).

Maturity, motivation, and rapport seemed to form a constellation in the eyes of the interviewers, with maturity correlating .60 and .59, respectively, with motivation and rapport. Motivation correlated with rapport .49 ($p < .0001$ for all three correlations). Age was not correlated with maturity ($r = .01$). Nonacademic achievement correlated with the constellation of maturity, rapport, and motivation .30, .40, and .24 respectively ($p < .0001$ for all three).

Low correlations noted in Table 1 may simply reflect the fact that students were selected from a narrow portion of the total applicant pool on that variable. Consequently, comparison of applicants rated highest and lowest on admissions variables might more effectively demonstrate the predictive value of the variables. Such a comparison is reported in Table 2. Students with highly enhancing levels of maturity, nonacademic achievement, and rapport were two to three times more likely to receive A internship letters than those without enhancing levels. (Two-by-two chi squares on these three variables with A and C letters versus highly enhancing and nonenhancing qualities yielded significant levels of less than .002 for all three.) High and low GPA and MCAT scores did not discriminate significantly between A and C letters.

Combining two or more relatively unrelated predictors might further discriminate between A and C letters. Because ratings of GPA and each of the personal characteristics showed little correlation, the investigators compared the combined effects of each of the personal characteristics and GPA as predictors of internship letters. Table 3 includes four three-by-three-by-three tables of internship letter categories tabulated by levels of GPA, one table each for maturity, nonacademic

TABLE 1

Correlations of Admissions Predictors with Internship Letters of 436 University of Missouri-Columbia Medical Students Entering 1966 to 1970

Factors	Spearman *r*
Maturity	.22*
Rapport	.22*
Nonacademic achievement	.18*
Motivation	.18*
Science Grade-Point Average	.08
Medical College Admission Test	
Verbal Ability	.04
Quantitative Ability	−.01
Science	−.03
Age	−.04

* $p < .0003$.

TABLE 2

Impact of 'High' and 'Low' Measures of Admissions Predictors on Internship Letters of 436 University of Missouri-Columbia Medical Students Entering 1966 to 1970

Category	No. of Students	Percentage of Letters*		
		A	B	C
Maturity				
Highly enhancing	145	33	42	25
Nonenhancing	78	10	38	51
Nonacademic achievement				
Highly enhancing	122	35	40	25
Nonenhancing	118	19	40	42
Motivation				
Highly enhancing	128	30	41	28
Nonenhancing	90	17	39	44
Rapport				
Highly enhancing	199	33	40	27
Nonenhancing	54	15	39	46
Total GPA				
3.5 or over	85	31	33	36
3.0 or under	180	19	42	39
Science MCAT				
600 or over	88	26	36	38
470 or under	53	21	50	30
Total Group	442	24	42	34

* Letter ratings: A = recommended with enthusiasm, B = strongly recommended, and C = recommended.

achievement, motivation for medicine, and rapport.

The following paragraphs describe the statistical analyses used to evaluate the relative contribution of GPA and each of the four personal characteristics (Table 3). In Table 3 GPA has been divided into low, medium, and high levels, and each of the four personal characteristics has been categorized as nonenhancing, en-

TABLE 3

The Relation of Internship Letter, Personal Characteristics, and Grade-Point Averages for 435 University of Missouri-Columbia Medical Students* Entering 1966 to 1970

Internship Letter Categories	Grade-Point Average								
	3.00 or under			3.01 to 3.49			3.50 or Over		
	Not Enhance	Medium Enhance	Highly Enhance	Not Enhance	Medium Enhance	Highly Enhance	Not Enhance	Medium Enhance	Enhance
Maturity									
Number of students	31	88	57	28	86	61	19	38	27
Percentage of letters									
A	3	18	25	25	21	31	0	37	56
B	35	42	47	36	48	48	47	32	19
C	61	40	28	39	31	21	53	32	26
Nonacademic achievement									
Number of students	50	82	44	46	81	48	22	32	30
Percentage of letters									
A	16	12	30	22	23	31	18	31	50
B	46	45	34	33	48	54	41	28	27
C	38	43	36	46	28	15	41	41	23
Motivation									
Number of students	36	89	51	36	88	51	18	40	26
Percentage of letters									
A	14	16	24	19	26	27	17	32	50
B	42	43	43	36	48	49	39	32	23
C	44	42	33	44	26	24	44	35	27
Rapport									
Number of students	22	84	68	20	67	86	12	26	45
Percentage of letters									
A	0	17	22	35	13	33	8	19	51
B	45	40	46	30	52	45	42	38	22
C	55	43	32	35	34	22	50	42	27

* For letter ratings, see Table 2 footnote.

hancing, or highly enhancing. In general, the higher the applicant was rated on GPA and personal characteristics, the higher was the internship letter rating.

The extremes of GPA and each of the personal characteristics clearly demonstrate the combined impact of these predictors. For example, only three percent of the applicants with low GPA and unenhancing levels of maturity received A letters, while 56 percent with GPA greater than 3.5 and markedly enhancing maturity received A letters.

To compare statistically the combined effects of GPA and maturity on internship letters, the first part of Table 3 was analyzed as though it were a three-by-nine table. The data yielded a chi-square of 47.55, $p < .00006$. Thus, GPA and maturity combined had a significant relationship with internship ratings. The next step was to test the contribution of maturity to the internship letter independent of GPA. When cells differing in GPA alone were added, a three-by-three table of internship letters by maturity ratings resulted. A chi-square (4 *df*) on this three-by-three table was 21.89, $p = .0002$. Thus, maturity alone had a significant relationship with internship ratings. The chi-square (4 *df*) of 21.89 for maturity when subtracted from the overall chi-square (16 *df*) produced a chi-square (12 *df*) of 25.66 ($p < .02$) (11). The last chi-square represented the impact of GPA on internship letters after maturity had been accounted for. Thus, GPA was a significant predictor of the internship letter ($p < .02$), though less so than maturity ($p < .0002$). A better, though less common, test in this situation is Kendall's tau for ordered contingency tables (12); it produced similar results, but with $p < .000003$ for the relationship of maturity and internship letters.

Analyses of the effect of the other three personal variables, nonacademic achievement, motivation, and rapport, to GPA and internship ratings were carried out in a similar manner.

For nonacademic achievement the overall chi-square (16 *df*) was 36.54, $p < .003$, indicating that the combined effect of GPA and ratings of nonacademic achievement had a significant relationship to internship ratings. Nonacademic achievement had a significant relationship with internship letter (chi-square [4 *df*] = 14.88, $p < .005$), as did GPA after subtracting nonacademic achievement from the combined effect (chi-square [12 *df*] = 21.66, $p < .05$). Kendall's tau produced a significance level of $p < .0003$ for the relationship of nonacademic achievement and internship letters.

For motivation the combined effect of GPA and motivation produced an overall chi-square (16 *df*) of 26.44, $p < .05$. Motivation did not have a significant relationship with internship letter by chi-square nor did GPA alone. Kendall's tau for motivation and internship letters yielded a significance level of .002.

The combined effect of rapport and GPA produced an overall chi-square (16 *df*) of 47.11, $p = .0001$. Rapport had a significant relationship with internship letter (chi-square [4 *df*] = 21.59, $p = .0002$) as did GPA after subtracting rapport from the combined effect (chi-square [12 *df*] = 25.52, $p < .02$). Kendall's tau ($p = .000007$) indicated that rapport and internship letters were significantly related.

Discussion

Ratings of personal traits obtained from admissions interviews had higher correlations with clinical success as measured by internship recommendation letters than did conventional cognitive variables such as undergraduate GPA and MCAT scores. Highest correlations occurred with the interviewer's judgment of the degree to which the candidate's maturity enhanced or detracted from his candidacy. Indeed, enhancing levels of maturity seemed almost a prerequisite for outstanding performance inasmuch as only 10 percent of students without enhancing levels of maturity received A letters. Further, students with highly enhancing levels of maturity, nonacademic achievement, or rapport were two to three times as likely to receive A letters as those without enhancing levels.

The fact that this study shows greater correlations of personal characteristics with clinical success than GPA with clinical success is not surprising. While admission committee members consciously paid great attention to the candidates' personal characteristics and discussed them at length in committee, most would agree subjectively that they placed greatest emphasis on GPA. Thus, the range of GPA was restricted. Further, it is likely that the clinical faculty in their assessment of students' clinical performance gave great weighting to the personal characteristics of students. Relationship with patients, willingness to work long hours, organizational skills, and judgment are commented upon in faculty evaluations at least as frequently as comments about basic intelligence and factual knowledge. These characteristics were perceived as highly important by the internship advisory committee.

This study shares some limitations with other studies. It is confined to a local population. Students were selected in a nonrandom fashion and represented a small proportion of the total applicant pool, thus restricting the range of predictors. There is limited knowledge of the reliability both of grade-point averages and of interview ratings. Despite these findings, the following observations can be made:

1. The significant correlations with clinical success suggest that further emphasis on the personal characteristics, including maturity and nonacademic achievement, would have enhanced the overall clinical abilities of members of these classes.

A decade has passed since the first of these five classes entered medical school. At that time the surplus of qualified applicants was not nearly as great as today. Consequently, 40 percent of accepted students in this study had college GPAs of less than 3.0. Today, these individuals, even taking into account grade-point inflation, would be discouraged from applying by premedical advisers and would not even be interviewed by most admissions committees. Nevertheless proportionally more of the students in this group obtained stronger internship recommendations than was the case for the group with low maturity (Table 2).

Each admission committee has no difficulty selecting the "winners," those achieving the highest scores in all measures. Problems lie in selecting the remaining candidates. The choice frequently is between applicants with high GPA but with few strong personal characteristics or candidates with high levels of maturity or nonacademic achievement but borderline GPA. The data (Table 4) would support a decision by committees to select applicants of reasonable GPA who have high personal characteristics rather than those of highest GPA who do not have enhancing personal characteristics.

2. Measures of nonacademic achievement have potential for screening candidates.

[redacted]

The screening process should function both to minimize the percentage of C letters and to maximize the percentage of A letters. Measures of both nonacademic achievement and GPA can be obtained

TABLE 4

Extremes of Grade-Point Average, Maturity, and Nonacademic Achievement as Predictors of Internship Letter*

GPA and Maturity					GPA and Nonacademic Achievement				
Admissions		Internship Letter			Admissions		Internship Letter		
GPA	Maturity	Percent		No.	GPA	Nonacademic Achievement	Percent		No.
		A	C				A	C	
High	High	56	26	27	High	High	50	23	30
Low	High	25	28	57	Low	High	30	36	44
High	Low	0	53	19	High	Low	18	41	22
Low	Low	3	61	31	Low	Low	16	38	50

* For letter ratings, see Table 2 footnote.

from medical school applications. The two measures have low correlations, and each probably does discriminate between A and C letters given an unrestricted range. Using both variables in the screening process has the potential to enhance the quality of the applicants brought to interview and thus to increase the likelihood of selecting outstanding clinical students.

For decades admissions committees have advocated selection of well-rounded candidates for medical school. This study provides validation for this approach. Overall, 31 percent of Missouri medical school graduates who in college had GPAs of greater than 3.5 received A level internship letters. If these individuals also had a highly enhancing rating on one or more of the personal characteristics, the likelihood of receiving an A letter increased to over 50 percent. On the other hand, if those with high GPA had nonenhancing levels on one or more of the personal characteristics, the percentage of A letters dropped to less than 20 percent (Table 3).

The value of these predictors needs to be reviewed in other populations. More precise definitions of these predictors may well lead to higher predictive values.

References

1. Brading, P. L. The Relationship Between Success in Medical School and Both Selected Academic and Nonacademic Prediction Factors. Unpublished doctoral dissertation, University of Southern California, 1971.
2. Gallagher, R. E. An Exploration of the Prediction of Academic Performance Through the Use of Adjusted Pre-medical Grades. Unpublished doctoral dissertation, Ohio State University, 1970.
3. Gottheil, E., and Michael, C. M. Predictor Variables Employed in Research on the Selection of Medical Students. *J. Med. Educ.*, 32:131–147, 1957.
4. Gough, H. S., Hall, W. B., and Harris, R. E. Evaluation of Performance in Medical Training. *J. Med. Educ.*, 39:679–692, 1964.
5. Johnson, D. G., and Hutchins, E. B. Doctor or Dropout. *J. Med. Educ.*, 41:1097–1269, 1966.
6. Richards, J. M., Jr., and Taylor, C. W. Predicting Academic Achievement in a College of Medicine from Grades, Test Scores, Interviews and Ratings. *Educ. Psychol. Meas.*, 21:987–994, 1961.
7. Johnson, D. G. A Multifactor Method of Evaluating Medical School Applicants. *J. Med. Educ.*, 37:656–665, 1962.
8. Mensh, I. N. Orientations of Social Values in Medical School Assessment. *Soc. Sci. Med.*, 3:339–348, 1970.
9. Clapp, T. T., and Reid, J. C. Institutional Selectivity as a Predictor of Applicant Selection and Success in Medical School. *J. Med. Educ.*, 51:850–852, 1976.
10. Rhoads, J. M., Gallemore, J. L., Jr., Gianturco, D. T., and Osterhout, S. Motivation, Medical School Admissions and Student Performance. *J. Med. Educ.*, 49:1119–1127, 1974.
11. Neyman, J. Contribution to the Theory of the χ^2 Test. In *Proceedings of Berkeley Symposium on Mathematical Statistics and Probability*, 1945–1946. Los Angeles: University of California Press, 1949.
12. Kendall, M. G. *Rank Correlation Methods.* (Third Edition.) London: Griffin, 1962.

EXHIBIT B

Predicting Supervisor, Peer, And Self Ratings of Intern Performance

Penelope Kegel-Flom, Ph.D.

The work of the physician is likely to be viewed somewhat differently by his colleagues, his patients, and by the physician himself. Each "viewer" may, in his turn, contribute a unique and meaningful perspective to the overall assessment of the physician's performance. In order to test, in part, this contention, supervisors of interns, intern peers, and interns themselves were asked to rate the performance of medical interns on four dimensions: knowledge of medical sciences, clinical effectiveness, interpersonal effectiveness, and overall competence/professional promise as a physician. Ratings from the three sources over the four dimensions were compared in order to determine elements of communality and/or difference in the ratings.

In addition, the author questioned the extent to which ratings of intern performance could be predicted from earlier measures of aptitude, achievement, and personality. To answer this question, an assessment was made of the efficacy of medical school admission data (premedical grades, Medical College Admission Test [MCAT] scores, and interview ratings), personality inventory measures,* and medical school grade-point average (GPA) and faculty ratings in forecasting the internship performance ratings. The interns rated were the entire male graduating class of 1968 of the University of California School of Medicine in San Francisco (N - 110). The interns, in residence at 54 medical facilities throughout the United States, were rated at the close of their internship year, June 1969.

Ratings of Interns

In general, supervisors rated intern performance higher than did the interns' peers or the interns themselves; supervisors' ratings were also more variable, ranging from a score of 1 to 12 on every dimension; interns, on the other hand, were reluctant to give very high or very low ratings to their peers or to themselves.

Ratings by supervisors on the four dimensions of performance were highly intercorrelated; correlations among self-ratings and among peer ratings were lower than those among supervisors' ratings, but they also tended to be more highly associated among themselves than with ratings from the other two sources. Cluster analysis of the rating correlation matrix indicated that supervisors, peers, and interns themselves had indeed rated performance differently and that these rater differences were greater than the raters' distinction between the various work dimensions, for example, between clinical effectiveness and knowledge of medical sciences. Three clusters emerged in the analysis, the first based upon supervisor ratings alone and the second and third based most heavily upon self and peer ratings, respectively. In subsequent analyses each intern was given three criterion "cluster" scores. Each predictor variable was then tested for its ability to forecast supervisor, self, and peer rating cluster scores.

Predicting Intern Ratings

[redacted] Premedical grades demonstrated a low positive correlation with ratings of intern per-

---

This paper is based on a doctoral dissertation completed at the Institute of Personality Assessment and Research, Berkeley, as part of the institute's continuing program of research in medical education.

Dr. Kegel-Flom is a research psychologist with the School of Optometry, University of California, Berkeley.

* Two inventories were administered during the first week of medical school: *The Strong Vocational Interest Blank* (SVIB), which measures the similarity of one's interests to the interests of persons successfully employed in various occupations, and *The California Psychological Inventory* (CPI), which assesses 18 positive personal traits, for example, responsibility, tolerance, and intellectual efficiency.

TABLE 1

CORRELATIONS BETWEEN ADMISSIONS, PERSONALITY, AND MEDICAL SCHOOL ACHIEVEMENT MEASURES AND RATINGS OF INTERN PERFORMANCE

	Ratings of Intern Performance		
	Super-vision	Self	Peer
MCAT			
Verbal ability	-.09	-.21	.22
Quantitative ability	-.06	.05	.17
General information	.06	-.16	.18
Science	.01	.07	.06
Premedical grades			
Cumulative	.12	.09	.12
Science	.11	.12	.08
Last two semesters	.15	.16	.03
Interview rating	.04	-.04	.12
California Psychological Inventory*			
Dominance	.01	.30†	.11
Flexibility	.03	.23‡	.12
Tolerance	.08	.03	.35†
Strong Vocational Interest Blank*			
Physician	.25†	.16	.07
Production manager	-.23‡	-.05	.13
Office worker	-.22‡	.23‡	.11
Credit manager	.22‡	.14	.16
Business teacher	-.22‡	-.07	-.04
Chamber commerce executive	-.20‡	-.02	-.17
Life insurance salesman	-.19	.26‡	.35†
Banker	.07	-.23‡	.14
Mortician	.06	-.25‡	-.21
Librarian	.13	.03	.34†
Purchasing agent	-.10	-.23‡	.28‡
Sales manager	-.16	-.20	.35†
Real estate sales	.08	-.18	.27‡
Pharmacist	.02	-.15	.33†
Medical school grades			
Year 1	.32†	.22	.39†
Year 2	.16	.32†	.22
Year 3	.29†	.38†	.13
Year 4	.41†	.30†	.26‡
Cumulative grade-point average	.32†	.46†	.35†
Faculty ratings years 3 and 4			
Academic performance	.37†	.29†	.29‡
Clinical performance	.47†	.35†	.26
Personal qualities	.43†	.31†	.31‡
Recommendation for internship	.44†	.35†	.30‡

* For these tests, only those scales which correlated significantly (Pearson Product Moment coefficient) with one or more criterion measure are listed. Tests were taken at medical school entry.

† $p < .01$; ‡ $p < .05$

formance; admission interview ratings were almost completely unrelated to ratings of intern performance. [redacted] premedical grades, or admission interview ratings.

A number of personality inventory measures correlated significantly with one or more of the ratings of intern performance (Table 1). Although different personality dimensions correlated best with supervisor, self, and peer ratings, respectively, interest scales relating to business and sales occupations tended to correlate negatively with all ratings. In other words, the interests of highly rated interns were decidedly unlike the interests of persons in those occupations. Interns given high ratings by supervisors had interests most like physicians; however, the physician interest scale was not significantly correlated with self or peer ratings of interns. Peer esteem was uniquely and most significantly related to the personality trait of tolerance. On the other hand, it was dominance and (lack of) flexibility which characterized interns who gave themselves the higher ratings.

Grades achieved during the four years of medical school and faculty ratings of student performance in the third and fourth years correlated well with ratings of intern performance (Table 1). Supervisors' ratings of intern performance were best predicted by earlier medical school faculty ratings. Self ratings were highly associated with both medical school grades and faculty ratings. Peer ratings tended to be less associated with medical school performance than were the ratings by supervisor and self.

In order to determine the best and most efficient combination of variables to predict ratings of intern performance, a step-wise multiple regression analysis was conducted using Norman's (1) double-split cross-validation method. The intern sample was divided into odd-numbered and even-numbered subsamples ($N_1 = N_2 = 55$), each predictor variable correlated with the intern performance ratings within subsamples and then multiple regression analysis conducted within each subsample. Equations developed to forecast ratings in Group I (odd-numbered subjects) were cross-validated on Group II (even-numbered subjects) and equations developed in Group II were cross-validated on Group I. The mean of the two cross-validation coefficients was taken as the best estimate of the degree to which ratings in the internship could be forecast from linear combinations of the predictor variables.

Supervisors' ratings were forecast best by combining the following variables: Medical school GPA, year one, medical school faculty rating on clinical performance in years three and four, and the physician interest scale. In other words, the interns rated most highly by their supervisors were those who did well in their medical science course work, excelled in clinical medicine, and possessed the interests of physicians.

The best combination of variables to predict self ratings was the cumulative medical school GPA, the dominance scale of the California Psychological Inventory (CPI), and (negatively) the flexibility scale of the CPI. In summary, interns who rated themselves highly tended to be forceful and self-confident but somewhat rigid persons who had achieved high levels in their medical school work.

Personality variables alone were the best predictors of ratings by peers; prediction was not enhanced by adding medical school performance measures to the regression equations. The best prediction of peer ratings was a combination of personality inventory variables which gave positive weights to tolerance and psychological mindedness and negative weights to scales measuring the interests of pharmacists and life insurance salesmen.

Conclusions

The findings of this study support the contention that the medical intern's performance is viewed somewhat differently by his supervisor, by his intern peers, and by the intern himself and that these differences may often be greater than differences between various aspects of work. When admission data, personality inventory measures, and medical school performance were considered together, it was found that supervisors' ratings of interns were best forecast by earlier medical school faculty ratings, self ratings were forecast best by a combination of personality inventory measures and medical school grades, and peer ratings were best predicted by a unique combination of personality inventory measures.

The fact that premedical grades, MCAT scores, and admission interview ratings failed to predict ratings of intern performance is entirely consistent with earlier studies of medical interns (2), psychiatric residents (3), and physicians in practice (4, 5). The performance of medical interns, like that of physicians in general, is apparently associated more with personal qualities and medical training than with relative levels of aptitude and undergraduate achievement.

References

1. Norman, W. T. Double-Split Cross-Validation: An Extension of Mosier's Design, Two Undesirable Alternatives, and Some Enigmatic Results. *J. Appl Psychol.*, 49:348–357, 1965.
2. Richards, J. M., Jr., Taylor, C. W., and Price, P. B. The Prediction of Medical Intern Performance. *J. Appl. Psychol.*, 46:142–146, 1962.
3. Holt, R. R., and Luborsky, L. *Personality Patterns of Psychiatrists* (Volumes 1 and 2). New York: Basic Books, 1958.
4. Price, P. B., Taylor, C. W., Richards, J. M., Jr., and Jacobsen, T. L. Measurement of Physician Performance. *J. Med. Educ.*, 39:203–210, 1964.
5. Howell, M. A. Personal Effectiveness of Physicians in a Federal Health Organization. *J. Appl. Psychol.*, 50:451–459, 1966.

Michigan Council of Deans And Cooperation in Medical Education

Ralph E. Lewis

A new spirit of cooperation has appeared in medical education in Michigan and is evidenced by the Michigan Medical Schools Council of Deans. Although the deans of some medical schools in other parts of the country get together periodically, the Michigan Medical Schools Council of Deans apparently is the only such organization with a full-time paid staff.

The council was organized in February 1973 with the appointment of a full-time executive director and the establishment of an office, although the deans of the four medical schools* had met since late 1970.

The deans, their associate deans, and other key staff members meet with the council director monthly. The council works cooperatively on problems and programs and is the entry point for requests to the medical schools for information, cooperative projects, and other joint efforts. The admissions officers, directors of continuing medical education, financial aid officers, curriculum officers, public information officers, student affairs officers, and directors of community medicine/primary care/family medicine meet periodically with the council director.

Accomplishments

Following are some of the accomplishments of the Council of Deans in its first two years of operation:

1. Cooperation with the Association of American Medical Colleges and the medical schools in California on a pilot matching program for medical school admissions. The three M.D. producing schools in Michigan participated in the study, which was concluded in conjunction with the schools' regular admission procedures for the 1974–1975 entering class. The results of the "match" were compared with

---

Mr. Lewis is executive director of the Michigan Medical Schools Council of Deans, Ann Arbor.

* University of Michigan Medical School, Michigan State University College of Human Medicine, Wayne State University School of Medicine, and Michigan State University College of Osteopathic Medicine.

EXHIBIT C

Economic and Racial Disadvantage as Reflected In Traditional Medical School Selection Factors

Bart Waldman

Abstract—Medical College Admission Test scores and undergraduate grade-point averages show that both lower income applicants in general and minority applicants in particular are at a competitive disadvantage in gaining admission to medical school. This study examines the extent to which income and race, viewed independently of one another, affect these selection factors. By analyzing data collected by the Association of American Medical Colleges on applicants to the 1976–77 entering class, factors associated with race were found to have a far more significant influence on applicants' credentials than factors associated with lower family income. An admissions policy based exclusively on GPAs and/or MCAT scores would proportionately exclude minority applicants from medical education but would have no similar effect on lower income white applicants.

Previous data published by the Association of American Medical Colleges (AAMC) have indicated that applicants to medical school from minority groups traditionally underrepresented in medicine generally present credentials reflecting a lower level of achievement on the Medical College Admission Test (MCAT) and in undergraduate grade-point averages (GPAs) than do their white counterparts (1, 2). Similar data have also shown that performance on these traditional measures bears a positive though lesser relationship to the applicant's level of parental income (1). However, [redacted]

This study was conducted in an attempt to answer these questions. Further, the investigator attempts to evaluate the answers in order to reach very general conclusions as to the comparative levels of economic and racial disadvantage inherent in the use of GPAs and MCAT scores as criteria for admission to medical school.

When reviewing these data, the reader should bear in mind that small discrepancies in GPAs and MCAT scores are not necessarily predictive of success in medical school. For example, comparative graduation rates for students with scores on the MCAT Science subtest of 500 to 549 and 600 to 649 were 93 percent and 95 percent, respectively, when analyzed for the more than 75,000

Mr. Waldman is special assistant to the president, Association of American Medical Colleges, Washington, D.C.

students who entered U.S. medical schools during 1949-58 and who were followed longitudinally through the fall of 1962 (3).

More recent reports indicate that retention rates are now even higher (4, 5). Ninety-five percent of all 1968-69 entrants had received the M.D. degree by 1973, and an additional 1 percent of the students were still in school, for a net attrition of only 4 percent. Although the retention rate for the underrepresented minorities was slightly lower than that for all entrants, 89 percent had already graduated by 1973, and 91 percent were expected to graduate eventually despite relatively modest undergraduate GPAs and MCAT scores (4). And first-year retention of black Americans has risen from 91 percent of 1971-72 entrants to 95 percent of 1974-75 entrants (5). Thus, the reader should not assume that lower performance on these measures necessarily implies that an individual is unqualified for the study of medicine.

Method

The data base for all displays and analyses is the AAMC Medical Student Information System records for applicants and acceptees to the 1976-77 entering class at U.S. medical schools. The two principal data collection sources for these records are the MCAT and applications submitted to the American Medical College Application Service (AMCAS). Data from these sources have been supplemented by routine admissions reports from each medical school.

The study utilizes the following variables:

1. Parental income. Those applicants reporting gross combined parental income of less than $10,000 are differentiated from those reporting parental income of $10,000 or more. Incomes reported are for the calendar year prior to taking the MCAT (in most cases for calendar year 1974).

2. Racial background. The racial descriptor is based on the applicant's self-description; although some data displayed pertain specifically to black Americans, all comparative analyses are between whites and "underrepresented minorities" (defined as black Americans, Mexican American, American Indians, and mainland Puerto Ricans).

3. Undergraduate college grade-point averages. The GPAs have been converted by AMCAS to a uniform system of grading based on a scale of 0.0 to 4.0.

4. Medical College Admission Test scores. Separate subtest scores for Verbal Ability, Quantitative Ability, General Information, and Science are reported.

The data presented reflect the total information available within the AAMC Medical Student Information System for the 42,155 applicants and the 15,774 acceptees to the 1976-77 first-year class. The analyses necessarily were limited to those individuals for whom GPAs, MCAT scores, racial identification, and parental income were reported. Grade-point averages were unavailable for 4,916 applicants (11.7 percent of all applicants) and for 1,298 acceptees (8.2 percent of all acceptees). Medical College Admission Test scores were unavailable for 869 applicants (2.1 percent) and for 190 acceptees (1.2 percent). Racial identification was not reported by 2,210 applicants (5.2 percent) and by 558 acceptees (3.5 percent). Parental income was not reported by 1,826 applicants (4.3 percent) and by 457 acceptees (2.9 percent). In addition, the mean GPAs reported in Tables 1 and 2 reflect only those applicants who filed at least one application through AMCAS and, therefore, exclude applicants who applied only to one or more of the 29 schools which did not participate in AMCAS in 1976.

TABLE 1

MEAN UNDERGRADUATE GRADE-POINT AVERAGES AND MEAN MEDICAL COLLEGE ADMISSION TEST SCORES OF APPLICANTS TO THE 1976–77 FIRST-YEAR U. S. MEDICAL SCHOOL CLASS, BY RACE AND PARENTAL INCOME

Academic Measure*	All Applicants		White		Underrepresented Minorities†		Black Americans	
	$10,000 or More	Less Than $10,000	$10,000 or More	Less Than $10,000	$10,000 or More	Less Than $10,000	$10,000 or More	Less Than $10,000
Undergraduate GPA								
No. with known GPAs	31,089	5,101	27,139	3,191	1,725	1,238	1,249	899
BCPM GPA‡								
Mean	3.25	3.07	3.29	3.22	2.64	2.64	2.54	2.58
S.D.	.48	.57	.44	.48	.58	.58	.54	.57
AO GPA§								
Mean	3.37	3.24	3.39	3.33	3.00	2.98	2.93	2.94
S.D.	.42	.48	.41	.44	.49	.49	.48	.49
Total GPA								
Mean	3.30	3.15	3.34	3.27	2.80	2.79	2.71	2.74
S.D.	.41	.48	.38	.42	.48	.49	.46	.47
MCAT subtest								
No. with MCAT scores	34,351	5,977	29,235	3,527	1,965	1,416	1,391	1,015
Verbal Ability								
Mean	545	510	553	548	466	438	450	424
S.D.	89	103	80	89	95	89	89	85
Quantitative Ability								
Mean	597	554	605	584	502	479	483	463
S.D.	95	104	83	90	95	92	90	85
General Information								
Mean	532	503	539	533	469	446	454	434
S.D.	74	84	66	76	74	68	66	63
Science								
Mean	579	535	589	571	480	456	439	440
S.D.	90	104	76	86	97	93	90	89

* Sixty-eight percent of grades or scores lie within one standard deviation of the mean (68 percent of all applicants with parental incomes of $10,000 or more had BCPM GPAs of between 2.77 and 3.73).

† American Indians, black Americans, Mexican Americans, and mainland Puerto Ricans.

‡ Courses in biology, chemistry, physics, and mathematics.

§ All other courses.

TABLE 2

MEAN UNDERGRADUATE GRADE-POINT AVERAGES AND MEAN MEDICAL COLLEGE ADMISSION TEST SCORES OF ACCEPTEES TO THE 1976–77 FIRST-YEAR U. S. MEDICAL SCHOOL CLASS, BY RACE AND PARENTAL INCOME

Academic Measure*	All Acceptees		White		Underrepresented Minorities†		Black Americans	
	$10,000 or More	Less Than $10,000	$10,000 or More	Less Than $10,000	$10,000 or More	Less Than $10,000	$10,000 or More	Less Than $10,000
Undergraduate GPA								
No. with known GPAs	12,549	1,655	11,021	1,006	757	462	556	317
BCPM GPA‡								
Mean	3.50	3.36	3.55	3.54	2.87	2.93	2.78	2.87
S.D.	.38	.49	.33	.34	.53	.52	.51	.52
AO GPA§								
Mean	3.54	3.43	3.56	3.54	3.18	3.14	3.13	3.11
S.D.	.35	.43	.33	.35	.45	.47	.45	.47
Total GPA								
Mean	3.52	3.39	3.55	3.54	3.00	3.02	2.93	2.98
S.D.	.33	.42	.29	.31	.45	.44	.43	.44
MCAT subtest								
No. with MCAT scores	13,475	1,842	11,680	1,080	814	505	597	343
Verbal Ability								
Mean	577	544	582	580	502	471	489	457
S.D.	81	98	78	83	91	87	88	83
Quantitative Ability								
Mean	638	600	645	631	542	525	529	512
S.D.	78	94	73	76	88	85	86	81
General Information								
Mean	552	526	557	555	490	467	478	456
S.D.	73	84	71	75	73	69	66	65
Science								
Mean	622	589	629	625	529	513	516	499
S.D.	67	85	62	64	81	78	76	77

* Sixty-eight percent of grades or scores lie within one standard deviation of the mean (68 percent of all acceptees with parental incomes of $10,000 or more had BCPM GPAs of between 3.12 and 3.88).

† American Indians, black Americans, Mexican Americans, and mainland Puerto Ricans.

‡ Courses in biology, chemistry, physics, and mathematics.

§ All other courses.

Otherwise, all data reflect the total pool of applicants to the 1976–77 first-year class who took the MCAT examination at any time.

Findings

[redacted](as reflected in GPA), [redacted] This relationship exists for both applicants and acceptees.

As shown graphically in Figures 1 through 3, the distributions of GPAs and of MCAT scores are very similar for lower income and higher income whites. Likewise, the distributions are very similar for lower income and higher income underrepresented minorities. Although the lower income applicants within the given racial grouping consistently perform slightly less well than their higher income counterparts, the difference between these two groups does not approach the magnitude of the disparity between white applicants and underrepresented minority applicants, even in the same income range.

These findings are supported by the mean data summarized in Tables 1 and 2. These displays show that the mean total GPAs of the higher income and lower income white applicants are very similar (3.34 and 3.27, respectively). The mean total GPAs of the higher income and lower income underrepresented mi-

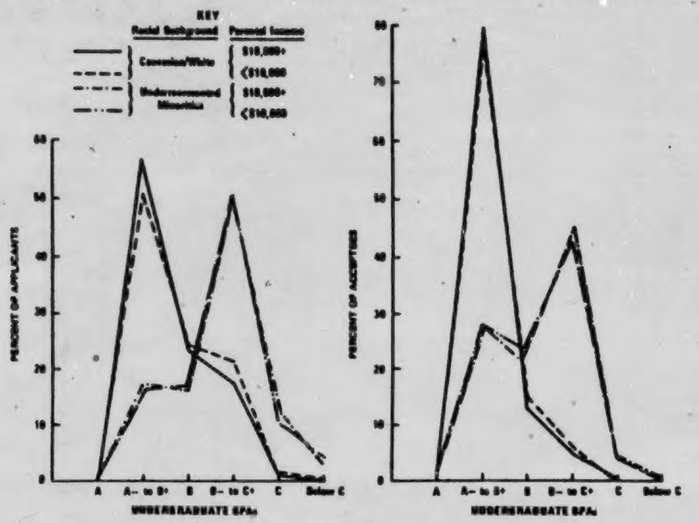


FIGURE 1
Undergraduate GPAs of applicants and acceptees to 1976–77 entering class of U.S. medical schools, by race and parental income.

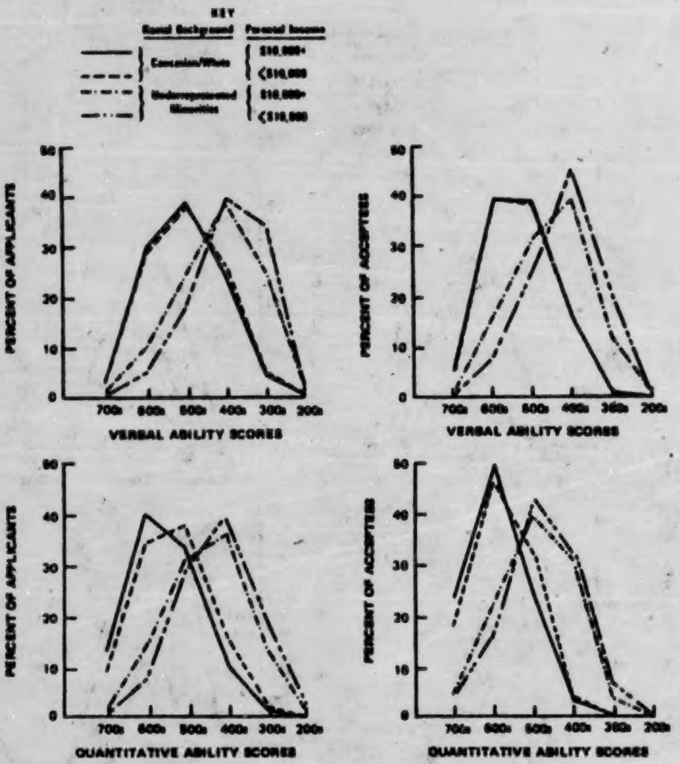


FIGURE 2

Verbal Ability and Quantitative Ability MCAT subtest scores of applicants and acceptees to 1976-77 entering class of U.S. medical schools, by race and parental income.

nority applicants are also essentially identical (2.80 and 2.79, respectively). However, the difference in mean GPAs between lower income white and lower income underrepresented minority applicants is .48, and the difference between higher income white and higher income underrepresented minority applicants is .54 (derived from Table 1).

Table 2 shows that the mean total GPAs of the higher income and lower income white acceptees to the 1976-77 first-year class are also essentially identical (3.55 and 3.54, respectively). Likewise, the mean total GPAs of the higher income and lower income underrepresented minority acceptees are virtually the same (3.00 and 3.02, respectively).

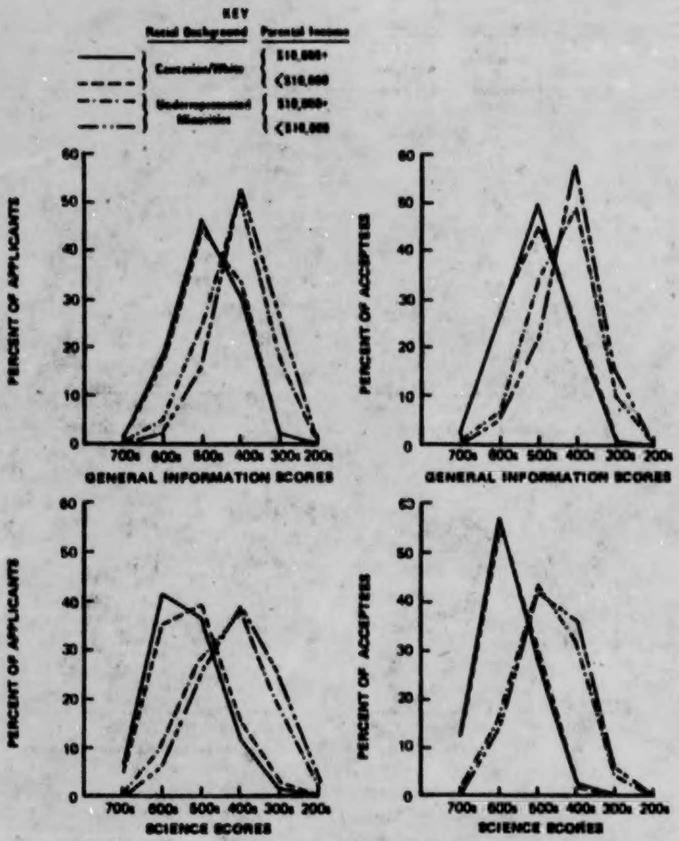


FIGURE 3
General Information and Science MCAT subtest scores of applicants and acceptees of 1976–77 entering class of U.S. Medical schools, by race and parental income.

However, the difference in mean GPAs between lower income white and lower income underrepresented minority acceptees is .52, and the difference between higher income white and higher income underrepresented minority acceptees is .55.

Similar results are found for the mean MCAT scores presented in Tables 1 and 2. For example, the mean MCAT Science subtest scores of the higher income and lower income white applicants are very similar (589 and 571, respectively). Likewise, there is only a 24-point difference between the mean MCAT Science subtest scores of the higher income and lower income underrepresented minority applicants. This compares with a standard deviation on that subtest of approximately 100 points. However, the difference in mean scores between lower income white and lower income underrepresented minority applicants is over 100 points, as is the comparable difference between higher income white and higher income underrepresented minority applicants (derived from Table 1).

Table 2 indicates that the mean MCAT Science subtest scores of the higher income and lower income white acceptees are also very similar (629 and 625, respectively). Likewise, the mean MCAT Science subtest scores of the higher income and lower income underrepresented minority acceptees are only 16 points apart. However, the difference in mean scores on the subtest between lower income white and lower income underrepresented minority acceptees is 112, and the difference between higher income white and higher income underrepresented minority acceptees is 100.

Table 1 also shows that 5,977 of the 40,328 applicants with known scores and parental incomes identified themselves in the under $10,000 income group. Of these, 3,527 identified themselves as white and 1,416 identified themselves with one of the underrepresented minority groups. The remaining 1,034 students either identified themselves with another ethnic/racial group or failed to answer the self-identification question. It is generally presumed that most of those failing to answer this question are white applicants.

Discussion

These findings demonstrate that an applicant's lower economic status alone results in a very slight, if any, competitive disadvantage because of lower performance in undergraduate work or on the MCAT. On the other hand, the variety of factors involved in minority racial status confers a far greater level of educational disadvantage which results in lower GPAs and MCAT scores and which is only slightly alleviated in the higher income minority group. These data do not permit an assessment of the social and cultural factors associated with race which underlie this educational disadvantage.

Figures 1-3 graphically demonstrate the degree of disparity in performance of minority and majority students on these traditional measures of academic qualification for medical school. They also show the high degree of congruity of performance of higher and lower income students within the same racial classification.

If medical schools had chosen to admit students based solely on one or more of these five traditional criteria, the 15,774 acceptances offered in 1976 would quickly have gone to individuals identified at the extreme high performance side of the curve. At this extreme, only a very small proportion of minority applicants could qualify.

However, admission based solely on these traditional criteria would result in only slight exclusion of lower income applicants, since their performance parallels that of the higher income applicants of the same racial group. Only to the extent that greater numbers of lower income applicants also belong to the minority population would lower income applicants be denied admission disproportionately. What seems evident, however, is that those students accepted from lower income families would also be overwhelmingly white.

For example, two traditional criteria often considered most influential by medical school admission committees are MCAT Science subtest scores and GPAs. If medical schools were to select students solely on the basis of MCAT Science subtest scores, the data indicate that a student would need a score of approximately 605 or better to qualify. This was derived by assuming that only whites and underrepresented minorities reporting racial self-description are applying for admission and by admitting only those with the highest scores. Of the 15,774 acceptances, 15,444 would be offered to whites and 330 to minority candidates. Of the total acceptances, 1,496 would be offered to students from the lower income group; 1,398 of these students would be white and 98 would be underrepresented minority group members. Thus, only 6.6 percent of the lower income students and only 2.1 percent of the total class would be underrepresented minorities. In comparison, underrepresented minorities would comprise 28.6 percent of the lower income applicants and 9.4 percent of the total applicants. Since this model ignores applicants and acceptees who are nonwhites not from underrepresented minorities or who failed to specify racial background (in reality, there were 5,167 applicants and 1,423 acceptees in 1976 who fell in one of these two categories), it tends to overestimate the degree of underrepresented minority representation which would be achieved.

If the medical schools followed the same admissions model but applied GPAs as the sole criterion instead of MCAT Science subtest scores, minority representation would climb to 448 and white acceptances would total 15,326.

TABLE 3

COMPOSITION OF FIRST-YEAR U. S. MEDICAL SCHOOL CLASS SELECTED ON THE BASIS OF VARIOUS MODELS*

Category	Total						White						Underrepresented Minority					
	All		$10,000 or More		Less Than $10,000		All		$10,000 or More		Less Than $10,000		All		$10,000 or More		Less Than $10,000	
	No.	Percent†	No.	Percent	No.	Percent	No.	Percent	No.	Percent	No.	Percent	No.	Percent	No.	Percent	No.	Percent
Applicants	36,143	100	31,200	86.3	4,943	13.7	32,762	90.7	29,235	80.9	3,527	9.8	3,381	9.4	1,965	5.4	1,416	3.9
Acceptees																		
MCAT model	15,774	100	14,278	90.5	1,496	9.5	15,444	97.9	14,046	89.0	1,398	8.9	330	2.1	232	1.5	98	0.6
GPA model	15,774	100	14,141	89.7	1,633	10.4	15,326	97.2	13,878	88.0	1,448	9.2	448	2.8	263	1.7	185	1.2

* Assuming applicants are limited to those whites and underrepresented minorities who actually applied for admission to the 1976–77 entering class and who reported racial self-description and parental income.

† All percentages reflect a comparison with the total figures in the first column.

The cutoff score for admission would be approximately a 3.38 GPA. Of the 15,774 acceptances, 1,633 would go to lower income applicants—1,448 to whites and 185 to underrepresented minorities. Thus, in this model 11.3 percent of the lower income acceptances and 2.8 percent of the total class would be underrepresented minorities. It must be noted, however, that this model makes no attempt to weight grades according to the quality or grading practices of the applicants' undergraduate institution.

In both of the above models the rate of acceptance of lower income applicants falls only slightly below their rate of application. Lower income applicants constitute 13.7 percent of the total pool of white and underrepresented minority applicants (10.8 percent of white applicants and 41.9 percent of minority applicants). In the MCAT Science subtest score model 1,496 lower income students were admitted. This amounts to 9.5 percent of all acceptances. In the GPA model, 1,633 lower income students (10.4 percent of all acceptances) were admitted. The data from these models are displayed in Table 3.

Although no school would be likely to apply a single criterion for admission, emphasis on these two criteria in the absence of any special admissions program would tend to produce results in the range predicted by these models. What seems apparent from these analyses is that while lower income white students generally would not suffer any severe disadvantage from the use of these traditional criteria, underrepresented minority applicants clearly would. Furthermore, the racial disadvantage reflected in these models and in Figures 1-3 appears to be based on factors which transcend purely economic considerations.

Thus, within a given ethnic/racial grouping, applicants' GPAs and MCAT scores vary only slightly according to parental income. But within a given income group applicants' GPAs and MCAT scores vary far more substantially according to race. From these observations it seems reasonable to conclude that lower income applicants as a class are not educationally disadvantaged in applying for admission to medical school, while applicants from underrepresented minority groups are disadvantaged by factors, extending beyond economics, which lead to a lower level of performance as reflected in GPAs and MCAT scores.

References

1. Gordon, T. L. *Descriptive Study of Medical School Applicants, 1975-76.* Washington, D. C.: Association of American Medical Colleges, 1977.
2. Johnson, D. G., Smith, V. C., Jr., and Tarnoff, S. L. Recruitment and Progress of Minority Medical School Entrants, 1970-72: A Comparative Study by the SNMA and the AAMC. *J. Med. Educ.*, 50:711-755, 1975.
3. Johnson, D. G., and Hutchins, E. B. Doctor or Dropout?: A Study of Medical Student Attrition. *J. Med. Educ.*, 41:1099-1269, 1966.
4. *Medical School Admission Requirements, 1978-79, U.S.A. and Canada.* Washington, D. C.: Association of American Medical Colleges, 1977.
5. Odegaard, C. E. *Minorities in Medicine: From Receptive Passivity to Positive Action, 1966-76.* New York: The Josiah Macy, Jr. Foundation, 1977.

Primary Care Education Trends In U.S. Medical Schools and Teaching Hospitals

Joseph J. Giacalone and James I. Hudson, M.D.

Abstract—In 1973 Schroeder and colleagues conducted a survey of U.S. medical schools in order to ascertain the nature of change in primary care education which had occurred by that date. In 1976 the authors of the present paper utilized the same survey instrument to document trend data over the three-year period. While conclusions from the 1973 survey pictured primary care education in a transitional state, the 1976 data suggest movement toward a general consolidation of efforts. The 1976 data indicate a continued emphasis on education in the ambulatory care area with no well-defined locus for coordinating institutionwide primary care training efforts, a leveling off of interest in health maintenance organizations (HMO) affiliations, no significant change in the number of affiliated programs for training nurse practitioners or physician's assistants, and a marked increase in the number of schools with affiliated generalist residency programs in internal medicine and pediatrics.

In the past decade new initiatives in primary care* education have been taken by virtually all the nation's medical schools and teaching hospitals. These programs vary in scope and in content. Almost universally, they have involved broader utilization of ambulatory care facilities for teaching; and in many instances they have resulted in major changes in affiliated residency programs. The changes so far initiated are likely to have serious implications with respect to future financing of medical education, particularly graduate medical education.

In 1973 Schroeder, Werner, and Piemme (2) conducted for the Association of American Medical Colleges (AAMC) a major survey of 116 U.S. medical schools, of which three were not involved in clinical teaching, in order to ascertain the nature of changes in primary care education which had occurred by that date. That survey supported the general impression that there had been intense activity in the development of primary care programs within these academic medical centers during the two to three years preceding the survey. What emerged was a picture of primary care education in transition with no clear pattern of optimum program development. The eventual shape of individual programs was projected as contingent both

---

Mr. Giacalone is staff associate and Dr. Hudson is director, Department of Health Services, Association of American Medical Colleges, Washington, D.C.

* The authors have adopted the Alpert and Charney (1) definition of primary care as first contact medicine which assumes longitudinal responsibility for the patient regardless of the presence or absence of disease and where the provider functions as an "integrationist" of health care for the patient.

Chairman **Perkins**. Dr. Bond, go ahead.

Dr. Bond is the director of the Learning Research and Development Center, the University of Pittsburgh.

STATEMENT OF DR. LLOYD BOND, RESEARCH ASSOCIATE, LEARNING RESEARCH AND DEVELOPMENT CENTER, UNIVERSITY OF PITTSBURGH

Dr. **Bond**. I would like to correct part of my record. That is not my correct title. I am a research associate there.

Mr. Chairman and members of the subcommittee, my name is Lloyd Bond. I am an assistant professor of psychology at the University of Pittsburgh and a research associate at the Learning Research and Development Center in Pittsburgh, Pa.

My academic training, teaching, and research are in the general area of research methodology, statistics, and psychometrics.

I appreciate the opportunity to testify before the subcommittee on proposed legislation regarding tests and their uses in higher education. Both as a psychometrician and as a citizen of color, I am both personally and professionally interested in any legislation regarding standardized tests and their impact on minority students.

At the start, may I say that I fully support many of the provisions of the bills under discussion and I support what I perceive to be their general intent. If these bills, in their final form, result in increased public knowledge and understanding of tests and how they are used, then the legislation will surely be in the public good.

I should like to summarize my perception of the rights of students and the public regarding testing and test use.

First, before they take any test which is subsequently used in decisions affecting their lives and careers, students should know what the test purports to measure.

Moreover, I would go even further and suggest that students are entitled to a comparable form of any test used for selection in order that they be familiar in a general way with the kinds of items they will encounter on the actual test.

Second, students and the public generally should have some assurance that tests are fair, that they contain unambiguous items, that they are reliable, and most importantly, that they validly predict the performance they are intended to predict.

Third, students have a basic right to know that test results are accurately scored and recorded.

Finally, students have a fundamental right to know how and to what extent test results will be used along with other information in making final selection decisions.

Much information relevant to the rights outlined above is already in the public domain. For example, a recent publication, "Taking the SAT," contains not only a complete sample SAT test, but also detailed discussions of the correct answer to each type of item.

Information regarding the relationship between test scores and the performance they are intended to predict is readily available from testmakers themselves, from college publications and from numerous other sources including the technical literature.

To state, however, that information is available to interested persons does not imply that everyone will avail themselves of the information. To the extent that any legislation increases the likelihood that test information will be more readily and systematically available to the public, then it serves a useful public function.

I believe many of the provisions of H.R. 4949 and H.R. 3564 will go far in answering the above concerns. I am deeply disturbed, however, with section 5 of the Weiss bill, H.R. 4949, which requires, upon request, public disclosure of test items after each administration of a test.

As a measurement specialist, I believe this portion of the bill is fundamentally misguided. It will in no way advance the rights and legitimate concerns of students and consumers, and I believe that it will eventually militate against black and other minority students.

Fair, unbiased achievement test items which validly predict future academic performance take months and even years to develop. Such items should be carefully guarded rather than revealed after each administration of a test.

The result of item disclosure will lead to the hasty and unprofessional construction of items which are likely to be ambiguous, more susceptible to cultural bias, and generally poorer in overall predictive value.

Further, I see no merit whatever in the belief that substantial instructional benefit will accrue from the examination of individual right and wrong answers. Admissions tests measure knowledge, competencies, and skills acquired over a relatively long period of time, both in and out of school. Increased performance on these tests result from sustained effort and quality instruction in academic areas. Examining individual items is a misleading, superficial, and meaningless exercise.

It is conceivable that the item disclosure provision will be a bonanza to the coaching schools since it will provide them with a ready source of items to use in their coaching sessions.

As such, it will be especially beneficial to those students able to pay for the facilities of these schools which is precisely the reverse of what is intended.

I should like to comment briefly on the current controversy surrounding the effectiveness of coaching schools in increasing performance on standardized tests. Part of the confusion on this issue results from the differing definitions of coaching.

Two- or three-hour cram courses, I am certain, will do little to increase performance on these tests. If, however, coaching is defined as sustained instruction in mathematics, grammar, vocabulary, and general reading skills, then, of course, such coaching will be effective. It is precisely developed competence in these areas that the test measures.

Some have claimed that item disclosure may aid in the detection of biased items on achievement tests. The technical literature is replete with attempts by content and subject matter specialists and by black and other minority educational researchers to search out and remove culturally-biased items from standardized achievement tests.

Overwhelmingly, the results of these attempts point to the same conclusion: Substantial differences in the achievement scores of middle class white students and various minorities remain even after presumably biased items are removed. In fact, many items which are not considered biased show larger mean differences between white and minority students than do items which are removed.

This result should not surprise anyone because the notion guiding the search for biased items is wrong. The search for such items seems predicated on the belief that the quality of education received by minority students is comparable to that of other students and that average performance differences therefore reflect some kind of cultural "extra-school" bias in the tests. And if we could get rid of this bias, everything would be OK. That is absurd.

Clearly, biased tests are undesirable, but I believe the continued research and concentration on this aspect of tests distracts both the scientific community and the public from the real issue here—namely, that minority education in this country is outrageously deficient.

A related belief is that somehow disclosure of test items will assuage the unequal quality of education in this country. This is simply not true. Quality items on standardized tests reflect in part the intolerably poor quality of education received in many public schools.

I do not believe, unlike my colleague, Dr. Jackson, that we should kill the messenger who brings the bad news. Rather, we should use the evidence to support further efforts to improve instruction and learning.

Mr. Chairman, other disadvantages of item disclosure have already been mentioned before this committee and I need not go into them in detail here. They include increased costs to students themselves, reduced services to students, less flexibility in terms of alternative testing dates, and less comparability across forms of the tests. These are important consequences for the students themselves. Any legislation should carefully weight these disadvantages against the presumed advantages of item disclosure.

There are other aspects of the proposed legislation which I believe could be changed for the better, including the requirement that unpublished and tentative research papers on standardized tests be made publicly available. Surely this portion of H.R. 4949, if it is enforceable at all, is counterproductive and may even increase public misunderstanding and suspicion regarding tests and test abuse.

A longstanding practice in academic circles has been that one's colleagues and others trained in a particular discipline should have the opportunity to critically review research papers before they go into the public domain. The proposed legislation flies in the face of this commendable tradition.

In conclusion, may I restate my firm agreement with legislative activity which increases public understanding and decreases the likelihood of test abuse. The importance of standardized achievement tests in the lives of students is great. Tests often carry more weight than they should.

But in our zeal to remedy overinterpretation and overreliance on test scores, we should be extremely careful to ensure that the effects of such legislation will not result in tests that are less useful and more susceptible to misuse.

I would like to comment briefly on some matters brought up earlier regarding predictive validity and differential predictive validity.

Since 1972, the University of Pittsburgh has started a very vigorous program to increase the number of black Ph. D.'s out of that program's psychology department. Prior to 1972, we had four black Ph. D.'s there. Since that time we have graduated over 20 in such areas as sociology, psychobiology, and mathematical psychology. We continue to use test information in the selection of those students.

There have been as many studies which were shown to not predict performance for a black student. I think I can cite twice as many which show positive correlations between black students' scores on academic tests and their performance in college.

I see nothing whatever wrong in asking a student to multiply two numbers. I see nothing whatever wrong with asking a student to define a word.

A recent study by the NAEP showed that 71 percent of black students in this country, high school students, could not do the following: A room is 12 feet long and has a total area of 96 feet. What is the width?

That is not a biased item and our continued concentration on this notion of bias as somehow artificially removing these absurd educational deficiencies engendered by our schools is a travesty.

Thank you, Mr. Chairman.

Chairman **Perkins**. Our next witness is Mr. Drexeler, president of the senior class, Great Neck, N.Y.

STATEMENT OF KENNETH DREXELER, PRESIDENT, STUDENT GOVERNMENT, NORTH SENIOR HIGH SCHOOL, GREAT NECK, N.Y.

Mr. **Drexeler**. I thank you for the opportunity to testify today. I am the president of the student organization which is the student government of the school, not just the senior class.

I think I can bring to you this morning a viewpoint which nobody else on the panel can, that of a high school student who has taken the PSAT, has taken the SAT's and who is studying to take the SAT again this Saturday and hopefully will be taking the LSAT in the future.

Our school has gotten involved in the question of the testing industry and the truth in testing bill primarily as a result of several incidents which occurred at our schools. There were errors made by the educational testing service. I described them in my testimony and have given them titles. They are not official titles but are used to give identification to the incidents.

First, the case of the missing paper and the duplicated test. In 1976 the Great Neck school district notified ETS that a test booklet of a SAT exam was missing. It subsequently developed that the missing booklet was taken by a Great Neck teacher who made copies and used it to give students practice in taking the SAT tests.

There was no question but that what the teacher did was illegal, but the test was used for purposes of coaching. Other examples are reported as missing and such tests should never be repeated. The teacher never contemplated, even as a remote possibility, that ETS would repeat the same test.

When ETS unaccountably did repeat exactly the same test only 2 years later, almost 50 Great Neck students, fewer than 5 of whom had even seen the earlier test, had their scores invalidated. ETS promptly brought a suit of copyright infringement against the teacher which was subsequently settled out of court.

No justice, however, was available to the 50 Great Neck students whose tests were invalidated due not to any fault of their own but to the gross inefficiency of ETS in repeating a test when a copy had been reported missing.

The October weekend fiasco: ETS, to its credit, is aware that some students observe Saturday as their Sabbath, and it gives such students the option of taking its exams on Sunday.

On Sunday, October 29, 1978, 13 Great Neck North students took the PSAT exams. Incredibly, ETS administered to them exactly the same test that was given to their classmates on Saturday, October 28.

The results: (1) The National Merit Scholarship Corp. disqualified the results of the 13 students who took the Sunday exam, and (2) to make amends, ETS asked the NMSC, which bases its scholarship awards on a formula based upon PSAT scorers, to use the regular SAT scores which the 13 students obtained the following spring in making the scholarship awards.

This has caused protests from others who claim that the scores on the regular SAT's are generally far better than those on the PSAT's, that the two are not comparable, so that in the words of a school newspaper editorial, "It's like comparing apples and oranges, or more appropriately, it's like comparing responsible organizations to irresponsible profitmaking groups such as the ETS. It just cannot be done."

How to demolish a student's college prospects: ETS permits a student to cancel an exam by filling out a cancellation report when he turns in his test paper. If he does so, ETS promises that it will not grade the exam nor make any report thereof.

One of the better students who had been rejected by some colleges later discovered that ETS had reported to all colleges to which he applied the results of an exam which he had canceled after completing fewer than half the questions.

There is no way to ascertain how substantially this ETS blunder affected the decisions of the colleges from which he was rejected. But how do you make restitution to this student? One must surely speculate as to how many times this has happened and gone undiscovered.

All three of these events occurred in one school with a population of 1,200 students in 1 year's time.

A recent article in the New Republic described a more serious and widespread incident:

> * * * There is no real recourse if ETS sends inaccurate information to a college about a student, because mostly likely the student will never find out about it. One such scandal became public during the 1975-76 academic year. Some of the LSAT

score reports sent to law schools contained an asterisk indicating that the applicant was an "unacknowledged repeater"—someone who had failed to indicate on the test registration form that he or she had taken the test before. Applicants were not told that overlooking the seemingly innocuous question about past tests risked incurring the deadly asterisk. Even worse, the ETS system accidentally labeled an unknown number of test-takers incorrectly as "unacknowledged repeaters." ETS eventually notified admissions deans to ignore the asterisk. But it made no effort to rectify the damage already done or to inform the victims of the error why they may have been rejected from law school.

Significantly, Ed Kiersh of the Village Voice reports that private citizens have brought more than 50 lawsuits against ETS or its clients during the past few years. Despite the numerous errors made in my school, I know of no family of a student affected by these errors who has filed such a lawsuit.

This certainly leaves us wondering how many blunders have actually been made but gone unpublicized and also unchallenged, primarily because the aggrieved students are dispersed throughout the country.

There is no central clearinghouse and therefore they have no communciation with each other as to errors that have affected other students. Overall, our unfavorable experience with ETS appears typical.

In protest of these regrettable situations, our student government unanimously passed a resolution strongly condemning the actions of ETS and requesting that an independent and impartial investigation be made of the testing organization.

Then, last May I organized a series of seminars at our school on the entire testing issue, including discussion of the New York truth in testing bill, the value of coaching schools, and the integrity of the testing industry.

The participants were Prof. Banesh Hoffman, author of "The Tyranny of Testing" and a longtime critic of standardized tests; Robert Moulthrop, director of the information division at ETS; and Allan Nairn, author of the Ralph Nader report on ETS. These seminars received coverage from the New York Times, Newsday, and CBS.

Perhaps the most serious and inequitable problem of all deals with the lack of security of the tests. The aforementioned incident in which a Great Neck teacher took a copy of an SAT exam is not by any means an isolated occurrence.

During the course of discussions I have had with teachers, both in my school and outside of my district, I am consistently told of the simplicity with which which one can take a test and that the practice is both frequent and widespread, especially among teachers who give preparation course privately.

Apparently, when the test booklets are counted at the conclusion of the exams, only a superficial check is made, often only the top sheets are counted. Thus, a teacher serving as a proctor will remove the inside of the booklet and merely return the cover sheet.

In preparation for the SAT exam which I will be taking this Saturday, I have just completed a $275 Stanley H. Kaplan coaching course. The most important feature of the course lies in that it allows students to enjoy prior familiarity with SAT-type questions and apparently actual SAT questions as well.

A relative of mine recently took a Stanley Kaplan course in preparation for the medical college admissions test (MCAT). Upon returning home from the exam the students in the course received telephone calls from the Kaplan organization asking them which questions they remembered. The Kaplan organization subsequently threw a party that Saturday night for all of the students in the course, during which the students were again grilled on their memory of questions. I have been told that this is a fairly common practice.

It is also thought by many that various test preparation centers will pay a group of people, often possessing so-called photographic memories, to sit through exams. Upon leaving the testing site they immediately write down the questions for use in prep courses.

Thus, at a recent session of my Stanley Kaplan course the instructor told us that virtually all of the questions either are taken verbatim from actual exams or are very similar. A report released earlier this year by the Federal Trade Commission affirms that taking a prep course can substantially increase a student's score.

The truth-in-testing bill requires the disclosure of test questions and this is important. For many years black students have scored consistently lower on standardized exams than white students who are on an equal level of scholastic ability by all other indicators. In most cases, these black students come from poorer homes and cannot afford to take expensive cram courses. With the adoption of the truth-in-testing bill in New York State, all students there will start out on an equal footing.

Obviously, the testing industry does not wish to comply with this New York law, hence many test producers have threatened to withdraw their exams from New York after the law takes effect on January 1, in order to circumvent this new law. This questionable tactic must not be permitted to work.

Another point which is in controversy is whether that would reduce the quality of the exams. The president of the AAMC says there is a limit to the number of relevant questions that can be used on the MCAT.

I think a Newsday editorial answered this best when they said:

> None of these claims is even marginally convincing. To argue that there's some limit on drawing up questions for prospective medical students suggest a dismally narrow view of professional education. Nor is it clear how preparing fresh tests periodically would compromise quality. Isn't it more likely that repeating the same questions would limit quality and invite misuse?

I think there is a fallacy in the way which many people are interpreting this bill. Dr. Howe mentioned this morning that testing is our greatest profession and that he doesn't know that we want to attack it.

The major point is that this bill does not attack testing. It merely insists that the testing industry be required to conform to certain standards of accountability and regulation.

I am speaking to you today not merely in my capacity as student government president, but also as a student who has already taken some of these exams, PSAT's, SAT's, and achievements, who will be taking the SAT again this Saturday, and who plans to be taking more tests in the next years, that is, LSAT's.

Though there is undoubtedly a need for some standardized testing, the testing industry must nevertheless be required to conform to certain standards of accountability and regulation in the interest of protecting the rights of all students and improving the selection process for higher institutions of learning.

Accordingly, I strongly urge the members of the House Subcommittee on Elementary, Secondary, and Vocation Education to support this national truth-in-testing bill.

Thank you.

[The information referred to follows:]

[From Newsday, July 23, 1979]

THE MEDICAL COLLEGES FLUNK THIS TEST

A reasonably assurate diagnosis can be made of the Association of American Medical Colleges' threat to retaliate against the state's new "truth in testing" law: It's symptomatic of acute arrogance.

The AAMC, joined by the American Dental Association, says it will cease to administer the Medical College Admission Test (MCAT) in New York after the first of next year. Virtually all medical schools in this country require student applicants to take that test; 5,000 New Yorkers sit for it every year.

How does the new law affect the medical examinations and others such as the College Boards? It merely permits students who have taken these tests to obtain their grading sheets and a list of correct answers so they can compare them and verify their marks.

What are the professional associations' objections? According to Dr. John A. D. Cooper, AAMC president, "there is a limit to the number of relevant questions that can be used on the MCAT." Dr. Thomas J. Ginley of the ADA says making the answers public would be "compromising the quality of the test." A spokesman insists that new questions would have to be developed every time the test is administered and that this would be impossible to do for some parts of the exam.

None of these claims is even marginally convincing. To argue that there's some limit on drawing up questions for prospective medical students suggests a dismally narrow view of professional education. Nor is it clear how preparing fresh tests periodically would compromise quality. Isn't it more likely that repeating the same questions would limit quality and invite misuse?

State education officials are meeting with the the associations to talk the matter over. They have a right to talk tough. The Educational Testing Service, a leading lobbyist against the "truth in testing" legislation in Albany, employed many of the arguments the AAMC and ADA are using, but it's not threatening now to suspend college tests or harass New York students. What do these two professional associations exist for, anyway—education or intimidation?

[The New York Times, July 24, 1979]

TRUTH IN TESTING

The intemperate responses of some academic testers to New York's truth-in-testing law are certain to damage their cause more than the law itself. The associations that administer admission tests for medical and dental schools have threatened to pull out of New York, and the Educational Testing Service, the nation's biggest testing enterprise, has said it might curtail services. Doubtless, the law will require more work on their part, and there will be added costs. But their complaints are overstated. In fact, the law is welcome; it's time to take the mystery out of college testing.

The law signed recently by Governor Carey applies to such examinations as the Scholastic Aptitude Test and the Law School Admission Test that colleges and graduate schools use to rank their applicants. As of Jan. 1, 1980, a copy of any test offered in New York must be filed with the State Department of Education. In addition, data on how a test was constructed and validated must be made public. What irks the testers most is a provision that allows anyone who takes an exam to receive a graded copy. Questions used for research but not counted in scores and specialized achievement tests are exempt.

The testers say the law means an exam given in New York cannot be used again because its contents are likely to be disseminated elsewhere. They contend, too, that the law will create problems in "equating" tests—insuring, for instance, that a 600 verbal score achieved in 1975 is comparable to a 600 in 1979. The testers therefore project drastic increases in cost to maintain the present level of service. And they're unhappy that the law goes into effect so quickly. This complaint may be justified. It takes months to prepare a test and a slight delay may be in order.

Testing services like ETS have been more open about themselves in recent years and the tests have improved; today, most are fairer to minorities than they were ten years ago. But the testers *still* haven't faced up to other exam limitations or to the frequent misuse of test results. Although testers caution against it, some schools still rely on scores as absolute indicators of ability. ETS, moreover, is surely exaggerating when it contends that the new law might drive up costs by as much as 50 percent and that even then it might not be able to uphold testing standards. Development costs are only a small part of of total costs passed on to the consumer. And some authorities have argued convincingly that at least four techniques already used by the industry would permit "equating" and validating to continue undisturbed—at little or no extra cost.

In any case, there is the matter of elementary fairness. These academic tests help to shape the course of people's lives—their schooling, their careers, the very sense of their own abilities. Students deserve to know how they are being rated and judged. There are freedom-of-information laws, truth-in-lending laws, truth-in-packaging laws. Why not truth in testing as well?

Chairman PERKINS. Mr. Andrews, you have a distinguished witness from your State and you may introduce him.

Mr. ANDREWS. Thank you, Mr. Chairman.

We have a witness from my State who has distinguished himself there and whom I would be privileged to introduce.

Dr. Cahow, we welcome you from Duke University.

STATEMENT OF DR. CLARK R. CAHOW, ASSOCIATE PROFESSOR, HISTORY, AND ASSISTANT PROVOST, DUKE UNIVERSITY, CONSORTIUM ON FINANCING HIGHER EDUCATION

Dr. CAHOW. I have been connected with the universities for the past 25 years. I have had the opportunity of seeing what has happened as a result of the improvement of test scores and the reporting of test scores.

I have come to the conclusion as the majority of the schools in the consortium on higher education have also concluded that there are some serious problems with the proposed legislation and I wish to speak on that issue.

We agree in effect with the previous testimony of Fred Hargadon representing the college board and Robert J. Solomon of the educational testing service. And while we do agree with them, we are not as concerned with ETA as we are with what might happen at the college and university level and with the students seeking admission in the colleges and universities across the Nation.

Editorial comments on the bills concerning the predictive validity of the standardized tests I think is sufficient and I will not comment any further on those matters.

I had the privilege of serving on the American Association of Law Schools Review Panel of the *Bakke* decision when it was issued, and I think if the *Bakke* decision spoke to any issue, it reaffirmed the fact that schools are guaranteed under the first and sixth amendments of the Constitution to admit students they choose to admit, to teach as they choose to teach, and I would hope that whatever comes out of this series of conferences and any legislation that might be forthcoming would keep this factor in mind.

The only concern of the court was that in the decision process there be a guarantee that there be no discriminatory actions taken against applicants. If this is a condition that is protected in the admissions process, the court has taken the position that standardized tests in fact can play and ought to play an important role in the decision process.

The literature of both the ACT and the college board speak to the fact that tests are constructed to assess students' general educational development and ability to compete in college level work, to require that a student demonstrate both problem solving and reasoning ability.

Both agencies readily state that the tests are good predictors of performance in college but they are not precise and should not be treated as they were. We take this into account as I think all responsible admissions officers take it into account.

For a number of reasons, no test test measures with precision. The reliability of the verbal section is considered 0.91 and 0.92 on the medical section. We think publication of test methods and results would damage this reliability.

Because we know all test scores have imprecision, we also know it is a generally accepted practice to interpret scores as ranges rather than as precise points. I am not trying to imply that test scores are worthless. They are not. On the contrary, we use test scores because we believe they measurably improve the validity of the admissions decision.

It has been demonstrated that the predictive validity of a high school record is better than the predictive ability of the SAT score alone. A combination has higher predictive value.

How tests are used in conjunction with other factors is a matter left to the individual school. No amount of disclosure of test information will alter the right of schools to use the tests as they see fit. We know that under this kind of condition test data can be misused. When only the SAT math score of two individuals differs 69 points on a 200 to 800 scale, is there any significant difference expressed in the ability of those two students?

Those of us in the admissions field know that this is the case and we know full well that the college board has cautioned against such misuse of tests.

Nonetheless, when a student with a 550 verbal is placed on the waiting list, one with a 540 verbal is rejected, and one with a 570 verbal is accepted, students raise the question and testing agencies outside the college board raise the question of the validity of the test.

A decision like this if it is not based on other factors is far from defensible. The common assumption that board scores are the primary determinant raises the level both of frustration and suspicion on the part of the applicant.

It also raises a level of frustration and extreme effort on the part of people in college admission offices when they are called upon to explain why it is that when board scores are put with other factors in the decision process, the school has admitted a student with a mid-range 800 score and has rejected a student in the 1100 range score, both from the same school.

This is a difficult procedure but one we accept as part of our responsibility. I actually know of no school that uses test scores in such an isolated and arbitrary fashion.

The test score by itself will not get a student in or keep a student out of a selected school. Schools that are not selective—and with the population drop as it is coming on us now in the next 10 years, more schools will not be selective—schools that are not selective, in those schools test scores play little or no role in the decision process.

Without tests we are convinced that it will be harder to discover and admit the student who for whatever reason of disadvantage is less known and less predictable without the test measure.

We have found this to be the case at Duke. I am sure it is the case in other schools in the consortium.

You may be interested in knowing, for instance, that the average SAT scores of entering freshmen in the 30 independent 4-year colleges of North Carolina ranged from 559 to 1,240. In only 4 of the 30 colleges did the average student have a score of over 1,000. The average scores of entering freshmen in the public senior institutions ranged from 588 to 1,063. This is information provided by the participating schools.

Only one university had average scores of over 1,000. At Duke, the entering class was composed of a wide variety of scores ranging from 700 to 1,560. The Duke range, incidentally, is similar to that of the 29 selective schools in the consortium.

Federal intervention in the matters of standardized testing when there is no perceived crisis will not change the way schools admit students. We believe that institutional reaction to variance in testing and the unintentional consequences resulting from hastily structured legislation will prove to be counterproductive to the needs to higher education and to the goals of students seeking admission to colleges.

We would urge caution in the consideration of the present legislation and a thorough review of the results stemming from the actions taken by New York and California.

We would further urge support of the testing services and the American Council on Education's proposals to undertake certain initiatives that will lead to:

> A mutual appreciation on the part of all concerned parties of the special relationships among the public's right to know, the university's need for educational autonomy, and the testing organization's obligation to sustain high professional standards in its work.

If I may make one or two comments about the previous testimony this morning, I would agree fully with Dr. Keller, that in all probability the testing services ought to provide data based upon linguistic groupings and report that data in norm groups by linguistic backgrounds.

I think this can be done. I think it can be done without legislation. I am concerned over the discussion that was held—and I may be muddying the water, I will try not to do so—over the issue of the way test scores are reported in both the SAT and the graduate record examination vis-a-vis men and women test-takers.

The scores are reported against men norms and women norms in both the verbal and quantitative areas in order that institutions reviewing those scores will not in fact discriminate against either men or women.

The norm is not a blind norm. It is a norm based upon men's performance and women's performance. This, I think, is what Dr. Keller was asking for in terms of reporting scores for minority students and it is one that I would support.

If there is no differential reporting of this kind for men and women, then schools will have less of a base for making nonbiased decisions.

I am inclined to agree with Mr. Shanker's testimony on the issues of test validity, cost, and the cram courses that we are convinced would spring up across the Nation.

The availability of test questions and answers will not necessarily solve the problem of providing a given individual the information he needs to improve his testing. Those questions and answers, coupled with an extensive and expensive, in all probability expensive, coaching program might help.

But it is the minority student, the economically disadvantaged student, who will not have the funds at the secondary level to secure that kind of coaching. Without that, for the students who are marginally able and who can afford the coaching and the possibility of skewing testing as a result, we see a real threat to the system that we work with now and that we understand in that it does allow us to evaluate a candidate's record at his school, is his nonqualifiable factors that we are looking for in our given student body and a validation of his ability to succeed at Duke and any other selective institutions. We see a danger posed to that process.

I believe as the queen of hearts, when she addressed Alice in Wonderland and when Alice asked how she should make the report and where she should begin and how long she should take, the queen said, start at the beginning and go to the end and stop.

Thank you, Mr. Chairman.

[The prepared statement of Clark Cahow follows:]

Prepared Statement of Clark R. Cahow, Associate Professor, History and Assistant Provost, Duke University

Mr. Chairman, my name is Clark Cahow. While I hold a position of Associate Professor of History at Duke University, my primary responsibilities as Assistant Provost, are in the areas of undergraduate admissions, university financial aid and records. I have over twenty-five years of experience in these two combined fields.

For the purposes of this hearing, I am privileged to represent a group of thirty colleges and universities known as the Consortium on Financing Higher Education (COFHE).

I thank you for this opportunity to talk with you today about the proposed legislation sponsored by Representative Gibbons in the Truth in Testing Act (H.R. 3564) and Representative Weiss in the Educational Testing Act of 1979 (H.R. 4949).

My presence here today is to record serious objections to the legislation pending before this Subcommittee. Duke University and the other colleges and universities represented in the Consortium believe that, as presently constituted, no proposed legislation could have more potential harmful results on the quality of the admissions process than these two bills. While we concur in the substance of the statements made before this Subcommittee by Robert J. Solomon of the Educational Testing Service and Fred A. Hargadon representing the College Board, our concern is less for ETS than it is for the institutions of higher learning that depend upon standardized testing as one aspect of their admissions procedure and for the students who seek admission to colleges and universities across the nation.

Testimony before this Subcommittee and editorial comments concerning the threat to the quality and predictability of standardized tests that is contained in these bills has been sufficient and does not require my further comment. I do wish to speak to the function of standardized tests in the admission process.

The Bakke decision, if it spoke to any issue, reaffirmed the fact that schools are guaranteed the right to decide how they will teach, what they will teach and who they will teach under the protection of the First and Sixth Amendments of the Constitution. As long as schools make their decisions, in the latter instance in a non-discriminatory manner, the use of standardized tests, in the opinion of the Court, represents a valid factor in the admissions decision process.

The literature provided by the Admissions Testing Program indicates that the Scholastic Aptitude Test "measures developed verbal and mathematical reasoning abilities that are related to successful performance in college." The literature prepared by the American College Testing Program indicates that the four academic subject matter tests "are constructed to assess (a) student's general educational development and ability to compete (in) college level work * * * (and) require that the student demonstrate both problem solving and reasoning ability," Both agencies readily state that "test scores are good predictors of performance in college (especially during the freshman year), but they are not perfectly precise * * * and should not be treated as though they were." All scores should, therefore, be used in conjunction with other factors such as high school grades and rank, and evidence of motivation—an intangible quality not measured by any test.

For a number of reasons, no test measures with perfect precision. The psychometric term used to express the degree of precision is called the reliability coefficient. This value which ranges from −1.00 to +1.00 indicates "the extent to which an individual would achieve the same score on repetition of a test"—that is the chances that John or Mary Doe will receive the same score from a second or third attempt of the SAT or other standardized test. Although not perfect, that is lacking a reliability coefficient +1.00, the reliability of the SAT-verbal sections is estimated to be approximately .91 and approximately .92 for the SAT-mathematical sections. The reliability coefficients of the ACT are somewhat lower than those of the SAT. They are between .74 and .77 for English usage; between .79 and .83 for Mathematics; between .71 and .73 for Social Studies; between .69 and .74 for Natural Sciences; and between .86 and .89 for the Composite. Because all test scores have a certain imprecision, it is a generally accepted practice to interpret scores as ranges instead of precise points.

I am by no means implying that standardized test scores are worthless. On the contrary, test scores measurably improve the validity of admissions decisions. It has been shown that the predictive validity of the high school record is higher than the validity of the SAF or ACT alone, and the combination of grades plus scores has higher validity than that of the high school record alone.

How tests are used in conjunction with other factors is a matter left to the individual school. No amount of disclosure of test informaton will alter the right of schools to use the tests as they see fit.[1] The scores of groups and individuals obviously can be misused. Examples of misuse might be acceptances or rejections based primarily on very similar SAT-combined scores such as 1230 and 1260 or ACT composite scores such as 19 and 20. The psychometric concept most appropriate here is the Standard Error of the Difference, which is roughly the square root of 2 times the Standard Error of Measurement. This value indicates "the normal variation to be expected between the scores of two people on the same test due to measurement error alone." Score differences of less than 1.5 times the Standard Error of the Difference have virtually no significance. This value varies by section of the test. To illustrate this concept, the Standard Error of the Difference for ACT subtests is approximately 4. For each composite score, it is approximately 2. Score differences of less than 1.5 times 4 equals 6 points on a particular subtest cannot be used to infer real differences in ability. For the ACT composite score, the Standard Error of the Difference is approximately 2. In practice, two students presenting ACT composite scores of 17 and 18, which is a difference of less than 2 points have, for all practical purposes the same ability as measured by the ACT. Also, it cannot be inferred that scores of 17 and 19, for example, on any of the ACT subtests, or any scores which differ by less than 4 points, signify real differences in ability. By comparison, for SAT-math scores, the Standard Error of the Difference is approximately 45. To use this value, our computation would be 1.5 times 46 which equals 69. This means that only when SAT-math scores of two individuals differ by more than 69 points on the 200 to 800 scale is there any significant difference in ability. For years, the College Board has cautioned against such misuse, but there is still the possibility of placing a 550-verbal scorer on the waiting list; rejecting a 540-

---

[1] I am indebted to Miss Thurletta Brown, Assistant Director of Admissions, Duke University, for interpretation of technical terms.

verbal scorer; while accepting a 560-verbal scorer. Such decisions are not defensible unless there are other factors in the application on which the decisions are based and which make the test score less than critical.

I know of no school that uses test scores in such an isolated and arbitrary fashion. Diane Ravitch, Teachers College, Columbia University, has put the issue in clear and succinct terms as it relates to the recently passed New York State law and as I believe it relates to the legislation under consideration before this Committee.

"Colleges ask their applicants to take standardized tests of verbal and mathematical skills because the scores are useful in comparing the aptitudes of students from all kinds of secondary schools—public, private, urban, suburban, rural. Colleges use the tests because they have some value as a predictor of future academic performance.

"College admissions officers are professionals. They know that standardized tests have limitations; they know that such tests reveal nothing of a student's imagination, creativity, motivation or character; they know too that some students freeze in test-taking situations.

"The scores are only one element in the student's record, which is made up of high school grades, rank in class, extra-curricular activities, teacher recommendations, work experience, and a personal interview."

A test score, by itself, will not get a student in or keep a student out of a selective school. At schools that are not selective, test scores play little or no role in the decision process. It will be harder to discover and admit the student, who, for whatever reason of disadvantage, is less known and less predictable without the test measure. Properly employed, test scores are a critical ingredient in the evaluation of a student's potential for success at a given school. The test scores provide an additional opportunity to discover candidate who will succeed who might otherwise have been overlooked. However, the clear assumption in the legislation is that colleges and universities use test scores to provide automatic cut-offs that preclude qualified students, particularly those that are "culturally deprived," from being admitted.

You may be interested in knowing that last year the average SAT scores of entering freshmen in the thirty independent four-year colleges of North Carolina ranged from 559 to 1240. In only four of the thirty colleges did the average student have a score of over 1000. The average scores of entering freshmen in the public senior institutions ranged from 588 to 1063. Only one university had average scores of over 1000. At Duke, the entering class was composed of a wide variety of scores, ranging from 700 to 1560. The Duke range, incidentally, is similar to that of the twenty-nine selective schools in the Consortium.

Many of the supporters of the Gibbons and Weiss proposals are seeking corrections for "abuses" that these proposals will not, in fact, correct. For example, students oppose the tests because of the power they believe the tests have in determining where they will be offered admission. If the tests have that power, giving the students access to all test questions will not change how the schools choose to use the test scores. Minorities oppose the tests and support the proposed legislation because minority students do less well, on the average, than other students. Again, giving students access to tests will not change how minority students perform on the tests. The only students who might do better are those who can afford to enroll in the commercial courses which coach students to take the tests. These courses will become even more popular once more sample questions become available from previous tests. The result will work against low-income, culturally deprived students, many of whom are minority students.

Federal intervention in the matters of standardized testing when there is no perceived crisis will not change the way schools admit students. We believe that institutional reaction to variance in testing and the unintentional consequences resulting from hastily structured legislation will prove to be counterproductive to the needs to higher education and to the goals of students seeking admission to colleges. We would urge caution in the consideration of the present legislation and a thorough review of the results stemming from the actions taken by New York and California. We would further urge support of the testing services and the American Council on Education's proposals to undertake certain initiatives that will lead to "a mutual appreciation on the part of all concerned parties of the special relationships among the public's right to know, the university's need for educational autonomy, and the testing organization's obligation to sustain high professional standards in its work."

Thank you.

Chairman **Perkins**. The last witness is Mr. William D. Geer, Jr., principal of the Francis W. Parker School in Chicago.

STATEMENT OF WILLIAM D. GEER, JR., PRINCIPAL, FRANCIS W. PARKER SCHOOL, CHICAGO, ILL.

Mr. **Geer**. Thank you, Mr. Chairman.

I am very pleased to be able to testify before this committee today and bring before you a matter of grave concern pertaining to the rights of students taking tests.

I have considerable experience with ETA and the College Board. Let me review my background to give you some idea of where I come from in this matter.

I am an English teacher and school administrator, presently a principal in a school in the city of Chicago. I was trained in English at Harvard and the University of Pennsylvania. I taught English at the Lawrenceville School in New Jersey from 1956 to 1965.

From 1965 to 1973 I was the principal of the Newton South High School in Newton, Mass. Since 1973 I have been the principal of Francis W. Parker School in Chicago.

I first worked for ETS and the College Board in the early sixties when I was employed to grade English achievement tests when they contained a section that had to be scored by readers rather than machines.

From 1963 to 1965 I graded English advanced placement exams. From 1963 to 1965 I worked on a special project in Trenton, N.J., and that was a project that involved special support, special education, and in an attempt to evaluate what particular elements on those tests were putting the black high school students at a tremendous disadvantage

In 1972 and 1973, while chairman of the Educational Policies Committee of the Massachusetts Principals' Association, I worked with CEEB to set up an experimental program to explore the use of achievement tests to qualify high school students to take university or college courses while still in high school.

I also served as a member of the council on entrance services of the College Board, a 3-year term running from 1972 through 1974. This is an advisory council which is given the responsibility for making recommendations to the College Board pertaining to the SAT and achievement tests.

Although I have had a number of criticisms of the College Board admissions testing program and the ETS about their testing programs, until recently I have always found that I could seek redress of the problems through the regular channels of review afforded by the College Board.

The matter I bring to you today I do not feel was successful in raising the issue through ETS or the College Board.

The matter I want to raise, I have written to all of the administrators of both ETS and the College Board and been involved in correspondence with them over the past 3 months.

The issue pertains very specifically to the question of the students' rights in a process of score review that ETS manages or has, I know, with SAT, and I assume with others. The score review is an attempt to use computers and other devices to ferret out and find students who have copied on tests. My knowledge of the system is based on telephone coversations and my specific experience with one student at my school that got involved in this.

Let me bring you into the matter and a matter of real concern to you, because I do think it's a clear question of a civil right of a student and probably thousands of other students who have been subjected to this process.

I became familiar with this practice in January when a student came to me for help after being accused of copying on an SAT exam. I have given to you copies of the entire correspondence between myself and ETS, responses from ETS, my letters, copies of my letters to trustees and two letters from Robert Solomon, who is vice president of ETS, which basically says, in my judgment, that ETS has every right to withhold students ’scores based on a system which is completely devoid of due process.

The student came to me on this whole question stating the student had been accused of cheating on the test. The student professed innocence, so I immediately began on the whole question of what could I do. The test was given in my school and I, as a reasonable person, said: "If ETS has accused you of this there must be some validity." I was handed the letter and could not tell what the basis of this decision was.

I had a lengthy series of discussions with the student and reviewed the student's record. The student had been in my school for 14 years. There had never been any question of the student's integrity; and she had a distinguished academic record and was one of the outstanding students in the school.

I also talked to the proctor of the test. The test session only had 24 students in it; and the proctor knew every single student in the room. The proctor said there were no irregularities. No student left the room. The desks were spaced properly at further distances, they say they have to be 5 feet.

I then entered into my lengthy telephone dialog with officials at ETS to determine the specifics and to ask that they review the matter. In short, over a period of 3 months I was unable to get a detailed statement of charges from ETS.

I finally received one in a month; but as I look at that there are so many things that are not there and there are obviously so many issues that are concerned that were not covered in that 3-page bill of particulars, and it was of no use to me to enter into any further discussion with ETS.

A review committee ruled that the student's academic record, testing history, and statement from the proctor were not sufficient to outweigh the evidence that they had. In early March the student took the SAT over again under controlled circumstances, and the score of the second testing confirmed the score that had been questioned, so we locked her in a room. She took the test again, and the score of the second test confirmed that the test they were questioning was valid.

It took ETS over a month to report the validated SAT scores to the colleges.

She took the second SAT in March, the first SAT in December, and no college received her validated SAT until April 24. In my judgment, ETS' actions almost certainly kept this student from going to a college of first choice. The student was a distinguished student and had a fine character record and has been able to go on to college; but in my judgment the question of where the student might have ended up as opposed to where the student did end up are two entirely different things.

From the last week in January until the third week in April a student's scores on an SAT were withheld because a computer in Princeton and a review committee using probability and statistics determined that "copying probably occurred." Neither the student nor I acting for the student had any effective way of questioning the allegation because we did not know the specific factors that led to the conclusion.

I understand the awesome responsibility of ETS to exercise all care and vigilance to detect cheating and copying on their tests, but given the importance of these tests in determining the future of a student, certainly from the testimony today, probably their system should be much more sophisticated and much more responsive than it is.

But given the importance of those tests, ETS must afford a student due process, and it's on that point that I specifically differ with Mr. Solomon. His contention is that ETS can withhold a score without any due process, that their review committee which is a group of professional employees of ETS has every right to withhold the score and all their responsibility is to inform the student that the score has been withheld.

I am concerned about the House bills, because it seems to me one of the objectives of particularly H.R. 4949 is to protect the rights of students and I am certainly not a constitutional lawyer, and I get into the question of wording bills and what would work and what wouldn't work.

It becomes very complicated, but I think that something should be done to make certain, no matter what is done with the rest of the truth-in-testing bill, to make certain that a student's rights in matters of score review are scrupulously protected.

I think what could happen; ETS simply could file with the Commissioner of Education a detailed account of what they did, that they had a review committee and here are the procedures, and that would meet all of the criteria of your bill.

There is no implicit requirement that a student's rights be protected. I also explored at some length with ETS the question of breach of contract; and that raises some very large issues. In my judgment, breach of contract is between ETS and the student. The student paid a fee to take the exam, to have the exam scored and to have those scores reported to colleges.

Clearly, in all of my discussions, ETS felt that their major obligation was to the colleges and universities. They would make statements like, we would be irresponsible to report to any college a score which we felt was in any way in question. Now, I was interpretive; but I think there are breach of contract issues there as well as constitutional issues, but I leave that to you, the lawmakers and the lawyers, to protect in some way, whether it's through a specific bill or whatever other way that you see as sufficient, the legitimate interests of students, the question of evidence, which I think this is the most important issue.

I would urge you to consider rewording article 3 specifically and maybe something could be done to protect the issue of evidence. So I would leave it with lawyers and others to look at those areas.

I don't have a specific wording to recommend, because that is not my field. The most troubling aspect of this whole matter to me has been the insensitivity of the officials at ETS to the impact of their practice upon students.

The letter notifying a student that ETS has concluded that copying has probably occurred is sent to the student and no one else so that the school doesn't receive it. The parents don't receive it. A student receives it. ETS defends this practice as a protection of a student's privacy.

[The prepared statement of William Geer, Jr., follows:]

PREPARED STATEMENT OF WILLIAM GEER, JR., PRINCIPAL, FRANCIS W. PARKER SCHOOL, CHICAGO, ILL.

I am very pleased to be able to testify before this committee today and bring before you a matter of grave concern pertaining to the rights of students taking tests.

Let me briefly review my professional background and experience with the CEEB Testing program before I survey the matter of my specific concern. I am an English teacher and a school administrator. I received an AB in English from Harvard University in 1956 and a MA in English from University of Pennsylvania in 1963. I taught English at the Lawrenceville School in New Jersey from 1956 to 1965. From 1965 to 1973 I was the Principal of the Newton South High School in Newton, Massachusetts. Since 1973 I have been the Principal of Francis W. Parker School in Chicago. I first worked for ETS and the College Board in the early sixties when I was employed to grade English Achievement tests when they contained a section that had to be scored by readers than machines. Following that, I graded English Advanced Placement machines. Following that, I graded English Advanced Placement exams for 3 years. From 1963 to 1965 I worked on a special project in Trenton, New Jersey, to try to determine some of the factors that kept successful black high school students from performing well on the SAT. In 1972 and 1973 while chairman of the Educational Policies Committee of the Massachusetts Principal's Association I worked with CEEB to set up an experimental program to explore the use of Achievement tests to qualify high school students to take univesity or college courses while still in high school. In 1972-1973 I also served as a member of the Council on Entrance Services of the College Board, the advisory council given the responsibility for making recommendation to the College Board on matters pertaining to the SAT and Achievement tests programs. Although I have had a number of criticisms of the CEEB Admissions Testing Program, until recently I have always found that I could seek redress of the problems through the regular channels of revies afforded by the CEEB.

This has not been the case in the matter I bring to your attention today. Although I have written to all of the trustees of both Educational Testing Service and the CEEB and to the administration of both organizations, neither organization has taken steps to protect students from the clear violation of their civil and constitutional rights present in ETS's process of score review.

I became familiar with this practice in January when a student came to me for help after being accused of copying on an SAT exam. I have included all of the correspondence between myself and ETS concerning the matter in the packet you have received and would be glad to answer any specific questions that you might have about the details, but let met briefly review the highlights. The student protested innocence, and a review of the student's record and general character as well as lengthy discussion with the proctor, who had taught the student, tended to support the student's position. I then entered into a lengthy phone dialogue with officials at ETS to determine the specifics and ask for a review. In short I was unable to get a detailed statement of the charges from ETS for almost a month. ETS ruled that the student's academic record, testing history and statements from the proctor were not sufficient to overweigh the evidence that they had. In early March the student took the SAT over under controlled circumstances and questioned score was confirmed. It took ETS over a month to report the SAT scores to the colleges, and in my judgement ETS's actions almost certainly kept the student from going to a college of first choice.

From the last week in January until the third week in April a student's scores on an SAT were withheld because a computer in Princeton and a review committee using probability and statistics determined that, "copying probably occurred." Neither the student or I acting for the student had any effective way of questioning the allegation because we did not know the specific factors that led to the conclusion. I understand the responsibility of ETS to exercise all care and vigilance to detect cheating and copying on their tests, but given the importance of these tests in determining the future of a student, ETS must afford students due process and the right of review before a score is withheld. Their present practice finds the student guilty, the score is withheld, and then the student, without knowing any of the specific facts that led ETS to come to its decision, is informed that the student may supply information to a review committee for consideration.

The student had contracted with ETS and CEEB to take the test, have the test scored and have the scores reported to colleges. Whether the issue rests in a breech of contract between the student and ETS, a direction that H.R. 4949 seems to explore, or whether the issue is really one of fundamental constitutional rights of denial of due process, I leave to the lawyers and law makers. There certainly was a callous denial of a student's rights when guilt was determined without any due process.

I have reviewed both H.R. 3564 and H.R. 4949 and am not certain that either or both of these bills if enacted into law would protect students against such practices. Certainly ETS and CEEB could meet all of the requirements under Section 3 of H.R. 4949 and continue their present practices. All that they would have to do is spell out the practice of score review under articles 6, 7 and 8 of Section 3. The only other part of H.R. 4949 that might pertain is Section 3, article 10C which I interpret to mean that the procedure for appeal or review of a test score by the test agency must be spelled out. I would urge you to consider rewording these articles of Section 3 so as insure that the civil and constitutional rights of the test taker be guaranteed. I feel that this can be done by requiring due process before a score is withheld and/or affording a student a detailed statement of the evidence and an opportunity to respond before a review committee makes its decision to withhold scores.

The most troubling aspect of this matter to me has been the insensitivity of the officials at ETS to the impact of their practice upon students. The letter notifying a student that ETS has concluded that copying has probably occurred is sent to the student and no one else. ETS defends this practice as a protection of a student's privacy, yet in my experience most students receiving such a letter in the middle of their senior year face the impossible situation of going to either their parents or school officials for help and being able to do nothing but protest their innocence because they have no idea of what the specific evidence is. Given the general anxiety about college admissions in the middle of the senior year, my guess is that most students, if they can, try to forget the letter. While talking to an official at ETS I was told that they are confident that the program of score review is effective because so few students ever respond to the initial letter from ETS. In fact ETS probably counts a nonresponse as an admission of guilt. How many shattered innocent students there are who received such letters in the six or seven years that ETS has engaged in this practice and could do nothing is a matter of conjecture, but having lived through the hell of one student improperly treated by this system, I can speak of the agony and confusion that it creates.

It is a legitimate interest of the Congress to protect the civil and constitutional rights of test takers, and I urge you to strengthen the present bills or draft a separage bill to do just that.

FRANCIS W. PARKER SCHOOL
330 Webster Avenue
Chicago, Illinois 60614

July 13, 1979

Dear

I am writing you to call your attention to a practice of Educational Testing Service connected with the SAT program that I judge arbitrary and irresponsible. The practice I question is the computer identification and review of major variations in the SAT scores of a student.

As a result of this practice a student at Francis W. Parker was informed, "that copying probably occurred," and that the scores for a December SAT were being held. I am enclosing copies of the letter that the student received, a letter I wrote supplying information and reasons why I thought the judgment unfair and the formal documentation that the "review" committee supplied me concerning their original judgment.

The matter was finally resolved when the student took the SAT over again and the score was confirmed, but the whole procedure had a crushing effect on the student. I object to the practice because:

> Although the student was not legally charged with cheating, the impact of the letter had the effect of a formal charge.
>
> Neither I nor the student could adequately respond to the "charges" because we were never informed in writing of their specific nature until after the "review" committee had met.

The scores for the SAT in question were not reported to all the colleges until April 19th yet the retake of the SAT was scored by March 14th.

The more I have learned over the past three months about this practice of ETS the more disturbed I become about it. ETS and the College Board are in fact serving the student who is paying for the administration, scoring and reporting of scores of the exam, yet in this matter it appears that ETS is willing to assume guilt without clear proof and is willing to sacrifice a student to the, "need to report reliable scores." Further the blind reliance on a computer model to detect "unreliable" scores represents bureaucratic arrogance of the highest order.

It seems to me that the concerns of ETS could be met by simply informing students, whose scores increase by more than a certain amount, that they should retake the test because most colleges will consider scores that have risen by that amount unreliable. ETS has created a considerable mechanism to seek, find, and deal with students whose scores fail to pass through the computer screen. In fact this system assesses guilt and then a student must prove innocence, a complete contradiction to our judicial system. The procedure of review has the superficial appearance of due process yet in fact there is none.

As a trustee you have the responsibility to investigate the practice of the review, and as it represents a misuse of the considerable public trust of both ETS and the College Board, see that it either ceases or is changed to remove its arbitrary and irresponsible elements.

Sincerely,

William D. Geer, Jr.

WDG:ew
Enclosures 3

EDUCATIONAL TESTING SERVICE PRINCETON, N.J. 08541

[illegible]
CABLE EDUCTESTSVC

February 1, 1979

Dear

Educational Testing Service (ETS) routinely checks the scores recorded for the Scholastic Aptitude Test (SAT). As a result of this general check, the Board of Review--a panel of senior staff members at ETS--is conducting a detailed review of the scores recorded in your name for the December 2, 1978 administration of the SAT.

This detailed review includes two comparisons. First, your current scores were compared with earlier scores and found to be quite different. Such large score differences are unusual but large differences alone are not sufficient reason for questioning scores. Second, your answer sheet was compared with others from the same test center. Your answers are in close agreement with those on another answer sheet. Such close agreement is unusual and indicates that copying probably occurred. The Board has decided that there is sufficient reason to question your scores because unusual results were found in both of these comparisons.

The enclosed pamphlet (ETS Procedures for Determining the Validity of Questionable Scores) explains details of all the options available to you in this review of your scores. Please read it carefully. In brief, these options allow you to:

> Request a retest to confirm your questioned scores.
> (Not available overseas.)
>
> Authorize cancellation of your questioned scores.
>
> Supply the Board with additonal information
> regarding the validity of any of your scores.

You will be informed of any further action once you reply. Meanwhile, your score report will read, "Absent or Scores Delayed"--a phrase used for all delayed scores. Colleges and others designated by you to receive

February 1, 1979

your scores will not be informed that a review is in progress. If you do not choose any of the options described in the pamphlet or if you do not reply, there will be no further action but your scores will not be reported. Please note that in your case this means the questioned scores will not be formally cancelled and the test fees will not be refunded as stated in the pamphlet. However, your previous scores or any future scores will be reported at your request.

Whenever the Board recommends against reporting a candidate's scores, there is one additional and special option available: You may request that the questioned scores and a statement of the Board's reservations about their validity be sent to any institution you choose, provided the institution is both willing to examine this information and to maintain its confidentiality.

Before you reply, you may wish to seek advice from your parents, counselor, or others. And if you have any questions regarding this matter, call me collect at 609-921-9000.

Sincerely yours,

(Mrs.) Shirley F. Kane-Orr
Secretary, Board of Review

SKO:cjb

Enclosure

FRANCIS W. PARKER SCHOOL
330 Webster Avenue
Chicago, Illinois 60614

WILLIAM D. GEER JR.
PRINCIPAL

February 7, 1979

Mrs. Shirley F. Kane-Orr
Secretary, Board of Review
Educational Testing Service
Princeton, N.J. 08541

Dear Mrs. Kane-Orr:

Enclosed is the material that I discussed with you yesterday concerning . As you can see from her transcripts, her grades particularly during the last two years in English and History courses have been excellent. Included in her transcript is our record of her achievement testing including the PSAT. Let me briefly review what I think is the valid information that should be presented to the Review Board. One, her verbal PSAT score was 45, one hundred points above her 35 SAT score which she took in March of 1978. She took Achievement Tests twice, once in May of '78 and once in November of '78, and as you can see from her transcripts, her English Achievement and Literature Achievement Scores have been 51 and 53 on both of those testings.

As I indicated to you on the phone, we do some testing at Francis Parker and we gave a complete battery of the California Test of Mental Maturity to in the third and sixth grade. In the third grade her scores were as follows:

Reading		Arithmetic		Lang.	
Voc.	4.7	Reas.	4.7	Mech.	4.7
Comp.	4.4	Fund.	4.9	Spel.	4.9
Total.	4.5	Total.	4.8	Total.	4.8

In addition to that we gave her the California Test of Mental Maturity in March of '73 in her sixth grade year and her scores were:

Reading		Mathematic		Lang.	
Voc.	9.4	Reas.	7.0	Mech.	10.3
Comp.	8.1	Fund.	3.3	Spel.	11.3

As I indicated to you also on the phone, has been working with the Stanley Kaplan program to improve SAT scores this fall. Although I haven't talked to the teacher at Stanley Kaplan, reflected to me that they had told her that she was consistently working in the high 500 on their practice test material. My understanding is that is going to have the person that she worked with at Stanley Kaplan prepare a letter for you.

My final concern has to do with your letter of March 1st to When we talked on the phone yesterday you indicated that you were not accusing of cheating on the SAT, but that your concern was merely one of reporting a true score to a college and insure a program which attempts, "to create a hedge against fate." Unfortunately in the second paragraph of your letter you state, "Your answers are in close agreement with those on another answer sheet. Such close agreement is unusual and indicates that copying probably occurred." The trauma that and her parents have gone through in responding to what they interpret as an accusation of cheating has been most difficult. As I indicated to you, college admissions situation was under review for March 1st and the delay in scores, I think, has already created a problem. As you commented on the phone, you felt that the letter was unfortunate and I would heartily concur. In my judgement no such letter should be sent to a student unless you have extremely clear evidence that the student has actually committed a wrong.

My understanding is that is mailing to you today a statement authorizing you to report her scores to the Francis Parker School and that you will also include a statement indicating the Board's reservation about their validity.

You indicated that you would send me an SAT at Francis Parker, and if chooses to take the SAT, I will administer that under controlled test conditions.

If there is any further information that you need from me please feel free to call me at the Francis W. Parker School.

Sincerely,

William D. Geer

William D. Geer, Jr.

WDG:ew
Enc.

EDUCATIONAL TESTING SERVICE PRINCETON, N.J. 08541

[illegible]
[illegible]

February 28, 1979

Mr. William D. Geer, Jr.
Principal
Francis W. Parker School
330 Webster Avenue
Chicago, Illinois 60614

Dear Mr. Geer:

Thank you for your letter of February 7, 1979 written in behalf of

This correspondence, which included grade transcript, and scores for the November and May 1978 Achievement Tests were given to the Board of Review for consideration.

As I informed you by telephone, the Board had reaffirmed its initial decision to require a retest in this instance and also directed that the material requested be released to you.

I have therefore enclosed both the initial Summary for the Board of Review and the resubmittal of February 14, 1979.

If I can be of further assistance, please telephone collect.

Sincerely,

(Mrs.) Shirley F. Kane-Orr
Secretary, Board of Review

SKO:cjz

Enclosures

SUMMARY FOR THE BOARD OF REVIEW

Probable Communication — January 31, 1979

Candidate: of Chicago, Il.

Scores Questioned by: 12/78 CBATP Computer Score Gain Check.

Candidate in question
CB
Score Record:

Date	Scores V	Scores M	Scores WE	Improvement V	Improvement M	Improvement WE	Improvement T
12/78	610	450	58	260	-60	9	200
3/78	350	510	49				

Candidate Whose Paper Corresponds:

CB
Score Record:

Date	Scores V	Scores M	Scores WE	Improvement
12/78	610	530	56	
3/78	630	500	54	

11/78 EN 570 LR 620 FR 550
5/78 EN 500 LR 580 FR 610

Documents in Evidence:

(1) Registrations and score reports.

(2) 12/78 answer sheets. 3/78 answer sheet for Candidate in question

(3) King's telephoned report that document for were written by the same person.

(4) Response analysis:

(1) Book numbers used:

Candidate in question used #139046
(name deleted) used #139047

(2) In the Verbal Sections, the correspond of incorrect responses is 64%.

(a) Candidate in question. 65 correct
14 incorrect
6 omits

(b) (name deleted) has 65 correct
12 incorrect
8 omits

January 31, 1979

(c) 9 incorrect answers common to both papers have identical responses. Index B - 168, result 6.

candidate in
(d) question has 4 answers correct that
does not have correct.

candidate in
(e) question has erased 10 items and changed to 9 correct and 1 omitted respons identical to selections.

(f) The longest run is 4 incorrect. Index E - 20, which is the total of incorrect and omitted responses on both answer sheets. The result is 6.

(3) In the Written English, there is no evidence of copying.

candidate in
(a) question has 45 correct
5 incorrect
0 omits

(b) has 43 correct
6 incorrect
1 omit

(4) In the Math Sections, 7 of 28 incorrect answers correspond - 25%.

candidate in
(a) question has 29 correct
28 incorrect
3 omits

(b) has 34 correct
15 incorrect
11 omits

candidate in
(c) question has 8 answers correct that
does not have correct.

(d) Erasures are not significant.

(5) In the Experimental Section, there is no evidence of copying.

candidate in
(a) question has 17 correct
(Math) 6 incorrect
2 omits

January 31, 1979

(b) has 36 correct
(Verbal) 5 incorrect
4 omits

Apparently copying was possible at the 12/78 SAT.

Recommendation: Complete security processing of the 12/78 SAT for

Shirley F. Kane-Orr

SFK/dm

CASE RESUBMITTED -

Date: February 14, 1979

Type/Program: Copy/12/78 SAT

Date of Previous Summary: January 31, 1979

Action Directed: Complete Security Processing

Additional Information Received:

1) Letter dated February 7 from Mr. William Geer, Principal of Francis Parker School.
2) Letter from authorizing release of all information from file to Mr. Geer.
3) Additional score report for indicating scores for 6 achievement tests that the Board had not originally considered.

Additional Action:

1) A retest has been sent to the school at Mr. Geer's request.

Recommendation: Inform that a retest is required and submit file to counsel regarding the release of information to Mr. Geer.

Shirley F. Kane-Orr

EDUCATIONAL TESTING SERVICE PRINCETON, N. J.

[illegible]
Executive Vice President

July 27, 1979

Mr. William D. Geer, Jr.
Principal
Francis W. Parker School
330 Webster Avenue
Chicago, Illinois 60614

Dear Mr. Geer:

Your letter of July 13 to Mr. Turnbull was referred to me because he is on leave for several months. We appreciate your taking the time to look into the case of one of your students and your general suggestions about the handling of questioned scores.

The basic problem, I think, is that no procedures are totally acceptable to all of the parties involved -- students, their schools, receiving institutions, testing centers and ETS. Our own view is that ETS has a responsibility for reasonable checks on the integrity of reported scores. In the process, we try to cause the least inconvenience or embarrassment to the student. We have no direct interest in identifying or labeling wrongdoing -- only in providing valid scores.

With no intent to minimize the reactions of the student about whom you wrote, allow me to review some of the issues that might not be apparent to someone outside of the system.

The computer is used in the process but only as the servant, not the master, of people who are examining the record carefully and making decisions. The computer does flag unusual score changes. The percentage of cases is very small (0.05%). However, the student is not contacted until the entire record has been studied and other facts raise a serious question about the validity of the score.

Your suggestion that ETS ask students to retake the test without suggesting that any irregularities exist other than a score gain sounds plausible and has been made before. We feel, however, that to suggest to the candidates that an admissions officer might question scores and urge retests based solely on a large score gain would be wrong. Such action would indeed endorse "blind reliance on a computing model" and properly draw criticism. Furthermore, what would we do if the student says that he does not choose to take our suggestion? Only if we acknowledge that we have a strong indication of copying or other irregularity do we have any right to withhold scores or otherwise inconvenience the student. We offer the student all reasonable options -- cancelling the score, retesting promptly under controlled conditions, supplying the ETS Board of Review with information that would alter the decision, or even authorizing the scores to be released to schools with a full record of the case. Note that unless the student chooses to release information, no public "charges" are leveled against him; he is simply told why ETS finds it necessary to inconvenience him.

You express concern that the student had no opportunity to "respond" to the Board of Review until after it had met. The initial meeting was for the sole purpose of deciding whether there was sufficient reason to question the scores or whether they should be released. As stated before, we do not consider a large score gain by itself to be sufficient reason to question scores. In this case we also found an unusual, close agreement with responses on another person's answer sheet; the serial numbers on the two answer sheets were consecutive. Thus, in the context of our considerable experience copying seemed probable.

It was at this point that the student was invited to respond with any information that would explain the large score gain and the close agreement with another answer sheet. The student was encouraged to consult others, to call us collect or to write to us. The student did turn to you for help and you were able to talk to our staff and, I trust, receive all the information you desired. The response that you prepared was carefully reviewed by the Board of Review at its second meeting, at which time the decision was reaffirmed to ask for a retest. We were pleased that the student elected the retest option and confirmed the validity of her scores.

Your concern about the delay between the retest and the reporting of the original scores is justified. I am informed that the delay resulted from clerical procedures that have already been changed. I have asked that these procedures be reexamined to ensure that delays will be minimal.

When we write to students about questioned scores, we include a leaflet describing the procedures and student options in detail. In case you did not have an opportunity to see the leaflet, I am enclosing one for your review.

The issues in dealing with questioned scores are complex and, as you must sense, something we have struggled with at length to find the most fair and effective procedures. Your interest is appreciated.

Sincerely,

Robert J. Solomon
Executive Vice President

cc: William W. Turnbull

FRANCIS W. PARKER SCHOOL
330 Webster Avenue
Chicago, Illinois 60614

ILLIAM D. GEER JR.
PRINCIPAL

August 23, 1979

Mr. Robert J. Solomon
Executive Vice President
Educational Testing Service
Princeton, New Jersey 08541

Dear Mr. Solomon:

Thank you so much for taking the time to write such a thorough and detailed response to my letter. Despite the thoroughness and detail of your response, there still are several questions which cause me grave concern.

My primary concern is the whole question which arises when ETS, for whatever reason, concludes that there is strong indication that copying or other irregularity has occurred. At this point I feel it imperative that ETS treat the matter in such a way that a student is afforded due process, and this would require that a student know the exact details of the charge and all of the information that has led ETS to draw such conclusions. As I indicated in my letter, I did not receive the detailed statement on the case until the first week of March, and I had been discussing the matter with officials at ETS since the end of January. Also, as I got further into the matter, I found that each time I discussed the matter with people at Princeton, another fact or issue would come to light. Lengthy phone calls with someone at your office in Princeton is no substitute for a clear and detailed statement of the causes and reasons why ETS feels that copying has occurred. I also strongly feel that there is no way that people sitting in Princeton, New Jersey, can come to an absolute conclusion that a student in Chicago has cheated simply by analyzing answer sheets and comparing patterns of right and wrong answers.

A further concern about the opportunity to respond comes from my feeling of uncertainty about exactly what the Review Board was considering. When I sent my letter, including background information on the student in question's academic record and general testing record, I had only the most general understanding of what the specific issues were that had led the Review Board to their conclusion that the student had probably copied. Unless a student, and those acting as advocates for the student, know the specifics, I see no way that the student can ever effectively respond to the conclusion that ETS has made that they have probably copied.

My final concern, after reviewing your letter, is that I am not clear as to whether your response constitutes a final statement of ETS on the matter, or whether ETS will conduct a more open and formal review of their procedures.

Sincerely,

William D. Geer, Jr.

WDG:ew
cc: William W. Turnbull

Chairman **Perkins**. I am going to interrupt you at this time so we will have an opportunity to ask Mr. Shanker questions.

He must leave at 1 o'clock, I understand. Mr. Shanker, I will start off.

You have been before this committee several times and you have had a wide experience in the educational area.

You referred in your testimony to the law which the New York Legislature recently enacted, and I understand a lot of the people who were backing that law today have some misgivings about the way that it has turned out.

However, that is just hearsay. But what do you think that we can learn from New York's experience? Assuming that the Weiss bill is enacted and implemented, what do you think about that situation?

Mr. **Shanker**. I think it is too early to say what we can learn. The law was just signed a matter of a few months ago. The impact is—with the exception of what the publishers have said they will do—the impact of that is still in the realm of conjecture, which is what all of us are, to some extent, engaged in here.

I would certainly think that, given the fact that such a law is there, that one of the things that might be worth doing is to wait a little bit longer here to see what the effect of that is, because right now with the exception of the threat of the publishers not to give some of the examinations there or release scores within the State, I don't think New York is in any better position to know right now what the impact of that is than we are here in discussing it.

We have got a series of conjectures and that is what they have got there, but we will know pretty soon and it's probably worth waiting here to take a look.

Chairman **Perkins**. What if this committee suggested a study, as suggested this morning by Dr. Howe, the first witness that testified; do you think that that might be helpful?

Mr. **Shanker**. I think continued studies of the issues raised here and what will happen in New York would be very worthwhile, yes.

Chairman **Perkins**. Do you feel that the enactment of this law would bring about higher costs, insofar as the testing is concerned?

Mr. **Shanker**. Yes; I believe—I am not an expert in this field—but there are people at this table who are, but I have not only heard them but I have met with others over a period of time and tried to listen to the different points of view.

I am convinced that it will increase the costs and will result in times on the examinations which are not as good.

It will result in a rush to action to create more and more and more tests and test items rather than better items, and that the consequences of the publication and release of the items will be negative rather than positive.

Chairman **Perkins**. Mr. Weiss?

Mr. **Weiss**. Thank you, Mr. Chairman.

First, let me express my personal welcome to Mr. Shanker for his presence and participation. We are associates. I am losing my voice.

Mr. **Shanker**. Associates and friends.

Mr. **Weiss**. Right.

I am sure, or if you don't, let me tell you that the area of greatest concern about the New York test and the New York law and its effect on tests refers to the small volume tests and the requirement, burden that would be imposed on making or publishing new tests in those instances.

Our legislation, I think, almost totally cures that problem, because we exclude tests which are given to a few of the people over the years. That bears cause to be reassured on that ground, but let me ask you, and I ask you these questions only because when we were in the process, my office with Shirley Chisholm and George Miller in preparing this legislation, and we tried to touch base with all elements of the educational community who would be affected by this legislation or would have a natural interest in it.

One of the groups that we touched base with was the American Federation of Teachers and its components at that time, and you correct me, let me ask the question first, but isn't it a fact that your union, the AFT, had in July of this year adopted a resolution at its convention supporting both the New York law and national efforts to reform teting procedures?

Mr. **Shanker**. Well, that is going a bit too far. Let me review that.

You are certainly right in saying that you had more than touched bases. You had met with and sought to find out what the views of both our State affiliate and national organizations were with respect to the issues involved here.

In New York State our affiliate did support, was not the prime mover, but did support the law which passed there. It has not changed its position. It is in the process now, as some of these consequences come out of examining its position, that I can only say while on the books, the position is there and it may end up being exactly the same as it was before; it may also be a few weeks from now: "Hey, we moved a little quickly and we wish we had given ourselves some breathing time," so I am reporting the policy in the State is there, and there is also a reconsideration.

The national resolution which was adopted refers to no specific piece of legislation, nor does it call for a specific. It does call for greater scrutiny. It is a one-sentence resolved with some whereases which point out we recognize there are problems and abuses. There is no question about that.

I indicated in my statement before that I am not here to take the position that nothing needs to be done. I am here to say that I am impressed with the complexity of the issues. I am convinced that moving very quickly on these is going to result in a lot of damage, and I would say that our organization, after there is enough either the same differences in our organization that exist on the other panels that you have had and this panel, I don't come here representing a half million people who have got unanimous opinions. I don't even have that many people that have looked at the issue.

I need some time. We have asked for and received a grant from the National Institute of Education to conduct a series of conferences for teachers around the country to bring people like the ones on this panel and bring teachers around this country, so this debate can go on all across the country, and we would like sometime to educate our people and involve them in this process.

We did the same thing a year ago on the question of reforming the tax base for education. We found most people in education around the country really were not aware of what the taxpayer and equity issues were. We ended up publishing a booklet and running a series of conferences and developed a good deal of quality leadership in this field.

I just am worried about conclusions at this early point. Now, you are right in your opening statement that so far as the only consequences on these special examples, they are given to smaller groups, but I don't think we have seen what the consequences of that legislation are. We won't until it plays itself out for a year or two.

Mr. **Weiss**. I just simply wanted the record to reflect our past involvement in coming to a conclusion, and also one further item I ought to mention for the record.

There was a representative of your organization that participated in the announcement of the introduction of this legislation, support of our efforts, and I assume that that was done with the authorization of you?

Mr. **Shanker**. I have not come here to complain. I haven't been consulted, and I will even say that you had strong reason to believe that you would have much more immediate support than you are getting from me at this moment.

Mr. **Weiss**. That is all that I wanted to establish.

Mr. **Shanker**. I will be very glad to acknowledge that.

Chairman **Perkins**. Mr. Goodling?

Mr. **Goodling**. Thank you, Mr. Chairman.

We have so little time that I don't believe I better ask all of the questions that come to my mind.

Mr. **Perkins**. Six or seven minutes.

Mr. **Goodling**. Just listening to the most recent conversation, there was a mention of excluding tests given to 5,000 individuals or less. I suppose Russian and Hebrew might qualify for such an exclusion. Beyond that, I can't imagine what other exams would be excluded, because when you talk about 5,000 nationwide, that is a drop in the bucket.

You answered the question I was going to ask you. I am a foot dragger, I suppose that is the way it is termed. So many times we have such well-meaning intentions and quickly write legislation, which we think is going to cure a number of problems. But instead, we sometimes do an injustice to the people we were trying to help. This has been mentioned many times today.

So I ask you the question that I have asked so many times:

Do you see such an emergency that we cannot wait until we see what happens with the California and the much stronger New York legislation, or the National Science Foundation study?

Is the emergency so great that we cannot wait and see before we do something wrong in attempting to cure the ills?

Mr. **Shanker**. Well, as I indicated before, I don't think the emergency is in the legislation that the people of the United States have this feeling that the testing is horrible and unfair and not valid and accurate. I don't think that the overwhelming majority, I think, the emergency is in another field.

I think that the people feel they are paying an awful lot for education and the scores are going down and the standards are loosening, and what we see at local and State levels all across the country is a way to figure out how we can tighten the standards, and I would say at the same time also use these results on some of these, the information on some of these tests to realize that we are not doing an adequate job in education, to try to figure out what we should be doing there.

My worry is that we have some instruments now. There are plenty of problems associated with the use of those instruments, and I think that Congressman Weiss and his associates who have introduced the legislation have certainly served a great purpose in bringing to light many of these problems. But I think that on balance, it is a lot better to try to improve an imperfect instrument than it is to open things up to the point where you may very well weaken the instrument, make it less sharp, less useful than it is at the present time.

You say you are a foot dragger. Well, I don't accept your evaluation of yourself. This is an unusual stance for me. I am usually here to say the following things ought to be done and I am convinced, but I do a lot of thinking and I do a lot of reading, and you have heard a lot of evidence, and there is sufficient amount that has been placed before you to show that the case has not been proven and that if you go ahead with this and the opponents are right, the consequences are pretty serious.

Mr. **Goodling**. Thank you, Mr. Shanker.

I have no other questions.

Chairman **Perkins**. Mr. Andrews?

Mr. **Andrews**. Thank you, Mr. Chairman.

If I may digress from the subject 1 minute and ask it very briefly of your friend from Kentucky, what happened to Dr. Ferguson who left the University of Greensboro and went to the University of Kentucky in about 1971? Is he still there?

Dr. **Clawson**. If he is there I don't know him.

Mr. **Andrews**. He was president of the University of Kentucky.

Dr. **Clawson**. Not in recent years. It is now President Singletary.

Mr. **Andrews**. I beg your pardon. It was Singletary.

Dr. **Clawson**. He is still president and doing very well.

Mr. **Andrews**. There are so many questions.

Chairman **Perkins**. All right; we will let your witness go.

Thank you very much.

Mr. **Shanker**. I would like to thank you for this opportunity and, if I may, I am going to run away, but one point that keeps recurring here is the belief that if somehow we throw the tests away that this is going to result in better admission standards insofar as the admissions of minorities are concerned.

I think when you eliminate standards of a core of objectivity, recognizing that this is only one of the pieces of legislation and if you use these tests on their own it's wrong and stupid. If you throw it away you are either left to local tests, and there is no evidence they are going to be any better, or you move over to various subjective standards.

Now, subjective standards may very well produce greater admission of minorities during a period of time when that is a popular cause. Then when that cause becomes less popular or unpopular, and all we have to do is see the swing of public opinion on various political issues and how quickly that happens, the abolition or the destruction of standards that move through different periods could lead to periods when there are greater admission of minorities than others when the subjective standards result in massive discrimination.

So one of the reasons for my support of standardized testing and objective testing and its improvement is a belief that in the long run minorities are advantaged by a rule of law rather than by a sway of opinion or popularity at any given time.

Thank you very much.

Mr. **Weiss**. May I comment very briefly on that?

The conclusion that you just made bothers me, because there is no effort on my part or anyone else's part supporting this legislation to——

Mr. **Shanekr**. I was responding to Dr. Jackson, not to you.

Mr. **Weiss** [continuing]. Throw away standardized tests. Quite to the contrary, I happen to believe that standardized testing has a place in the educational field.

My concern is the slipshod way which it is operated now, both as to validity and as to security which has destroyed the credibility of those tests. I want to undertake a system of opening it up, so it will regain credibility and sometimes there is a confusion as to whether in fact who is in fact more likely to secure acceptance of valid tests, those who want to reform the process or those who are willing to stick with the process, which everybody admits is terribly foolish.

Chairman **Perkins**. Let Mr. Shanker respond.

Mr. **Goodling**. I want to respond to that in relationship to this conversation.

I would like to read a sentence from Dr. Bond's testimony, because I think this is the key.

> The result of item disclosure will be to the hasty and unprofessional construction of items which are likely to be ambiguous, more susceptible to cultural bias and generally poor in overall predictive value.

I believe that is what the legislation as it is presently written will do and, therefore, I have to agree with most of what you have to say.

Chairman **Perkins**. All right. Let the gentleman finish his testimony now.

Mr. **Geer**. Let me pick up the point again.

The point was that there are clear civil rights issues as a principle; and as a person who personally knew this student I was most appalled at the impact of the letter and the action taken by ETS on the student.

The letter was received in the middle of the senior year. There was really no way for the student to question or to have any idea of why this decision was made. Thinking about that, the question of what does a student in the middle of their senior year do if they receive such a letter?

Do you go to your parents?

Do you go to your school guidance counselor, your school principal and say, "I received this letter; what do I do?"

Obviously, a student alone has little or no access or ability to question the decision or to even ask for an effective review.

Given the general anxiety about college admissions in the middle of the senior year, my guess is that most students, if they can, try to forget the letter. I was told that their program of score review is effective, because so few students ever respond to the initial letter from ETS and in those conversations my understanding is that this practice has been going on for 6 or 7 years, and probably involved thousands and thousands of students.

ETS probably counts a nonresponse from any student as an admission of guilt, and in their evaluation of their program they assume every time they send a letter like that out and receive no response from the student, the student is admitting that the student is guilty.

How many shattered innocent students are there who receive such letters in the 6 or 7 years? ETS has engaged in this practice and could do nothing about it is a matter of conjecture; but having lived through the hell of one student improperly treated by the system, I can speak of the agony and confusion that it creates. It is a legitimate interest to the Congress to protect the civil and constitutional rights of test takers.

I would urge you to clean the present bill or draft separate bills to do just that to deal with the right of the civil rights of students in the matter of the specific review of tests and the accusations of copying or cheating or irregularities.

Chairman PERKINS. Let me congratulate all of you. Now, I will try to hold our questions to 5 minutes the first time around.

I would like to ask the distinguished doctor from my own State a question, Dr. Clawson.

Before the national medical school admissions test was developed, how did medical schools choose students, and do you see any danger in going back to those practices if the national test must be made public after it is given?

Dr. CLAWSON. I will try, because each college obviously has different admissions standards and have had for many years, but I think the fact that American medicine has underrepresented groups, and it speaks for the fact that the previous practice was not the best practice and for the most part it was a matter of your knowing an alumnus or your daddy went to the school or you had some contacts with people on the admissions committee, and the process was less objective than it is with the MCAT.

I hold no grief. The MCAT is still a better tool. It is a better tool now than it was in 1975, and my concern is in listening to people discuss it as a bad tool, what is going to take its place?

I am very much concerned that what would take its place is not a better test, but going back to more subjective ways of admitting students which then there are inherent biases within the admissions committee, that I think would be damaging.

The comment by Mr. Gastfriend about the role of the MCAT, I believe, he testified most on MCAT as I did, stating the MCAT position, stating this was a great and good predictor of practice.

I have been with that organization in one way or another for a long time, sitting on its council of deans for 4 years, and I don't believe that that is the perception that most of us have. It is only a tool. I am surprised, hearing the amount of weight that admissions committees put on the standardized testing.

I don't hear that much weight being put on it, unless it seems to highlight something that we are missing from the other criteria for admission, and so my concern is that without an instrument that we have reliability or validity, or feel we do, and we will go back to some of the previous practices of either giving an examination locally in the school or using some of the other criteria, so that is my basic concern in this.

I feel that we have in the voluntary sector put a great deal of money into trying to make the examination better. We do publish a booklet that is available to everybody as to what the exam looks like, how the questions are structured; and we have answered many of the things voluntarily.

The question that was raised to my left here about an editorial that appeared, heavens, there are plenty of questions. Sure; we can construct questions forever. You can construct questions in chemistry that will test the mind of a Ph. D. That is not our desire in constructing this examination.

That is why the pool of questions is very limited and those writing the editorials did not recognize that we do not want to force students to take 5 or 10 years preparing themselves to get into a medical school by the way in which the test is constructed.

We have purposely limited the pool of questions to a very small number.

Chairman PERKINS. Mr. Goodling?

Mr. GOODLING. Thank you, Mr. Chairman.

I don't have enough time to ask all the questions I would like to ask. I am scheduled to meet with a group of my constituents.

Dr. Jackson, I would like you to send to me a response to a couple of comments that I heard you make during your testimony, which I interpreted as pretty much antistandardized testimony. You said at the end of your testimony that at Howard University the standards are going up and up.

Would you respond to me in writing? I got the impression that you also were inferring that there were some other tests involved, but the standards are going up and up at Howard. Can you explain as measured by what or by whom and are they going up in relationship to some testing tool that has been developed there?

Then, is there any possibility of teaching the testing tools or anything of that nature? It will take too long to respond to these questions now, but if you could respond in writing. I would be certainly interested. If you have a better tool I would want to get behind a better tool.

David, you present an interesting picture as I was thinking of pre-World War II. I don't know your background and family situation, but prior to World War II you wouldn't have gotten into the institution you attended unless you had previously attended some prestigious prep school or someone in your family had the proper connections.

However, you also present the interesting picture; in all probability prior to pre-World War II that institution wasn't prestigious enough to get you into a lot of medical schools, and that is one of the concerns I have.

Now, there is no question when you talk about feedback. Students should have feedback. However, if that means releasing these tests, then I have a problem with that, because you go on to say equal opportunity to prepare. If I thought there was any way we on the Federal level could provide equal opportunity for all young people to prepare for tests, I would certainly get 100 percent behind it. But to do that, I would have to make sure that they have an equal opportunity for all the cultural experiences as they are growing up that many do not have.

I would have to make sure that they have an equal opportunity to have a stable family situation. If I provide them test releases, then what I am doing to them, I think, is making it even more difficult for them, because then all of those agencies that want to make a quick buck get involved in this business of tutoring, et cetera, and they can afford it.

Mr. Gastfriend. I will respond to that, because I assume that you are asking me a question. I am here to give testimony rather than hear it.

Mr. Goodling. Yes, but you are also here without too much experience, and so we want to share our experiences with each other, because it will make us both better people.

Mr. Gastfriend. I am all for good people.

The question about what happens with the releases of the scores can be answered pretty easily by looking at a college bookstore in which you find books with compilations of questions, some of which have been gotten through the means of the Kaplan courses and others made up by educators, people who have even been involved in the test preparation industry; that is, working with ETS, the MCATR, and things like that who later go and write preparatory books.

You find that students do go out and purchase those books and review these things, but that is not the same thing as finding out why you scored a 400 or a 520 when you thought you did well, and it's not the same as going through and seeing where you flawed.

Looking at the exact questions that you answered, and you may feel that that is not significant, but I will tell you in the minds of students it is significant, and they may be inexperienced, but at present these are people who do not have a representative voice in the powers that produce, utilize and make policy on these tests, and that those students could not obtain such a voice through conventional means such as receiving appointments to those groups, but one way in which they could have a sense of ownership over the contract that they are making with the schools and testing services is just to know what other people think they are doing.

It's curious to me that some of the members of the committee who have asked some speakers to respond have not asked or not commented on the factual accounts of problems in the score reporting and the security problems that we have heard factual accounts of, and we have not heard you question further the data of the flaws in the predictive ability of the tests.

I sense a bias.

Mr. Goodling. You are going to take all of my 5 minutes in your response.

Mr. **Gastfriend**. Perhaps you would be interested to respond in writing to me.

Mr. **Goodling**. I am sure Mr. Geer didn't expect us to throw out the baby with the bath water because of several isolated incidents. You have to look at the overall picture.

Dr. Bond, may I merely say that I thought your testimony was outstanding. If I had any influence in the White House at all I wouldn't be so scared about a Department of Education, because I would make you the new Secretary of the Department of Education. Then I know that we would work at the problems we have in education.

Kenny, I would only say I hope if you have the profit-making information about ETS that you get it to IRS. I am sure IRS would be very pleased to have any information about ETS as a profit-making organization.

I think you should make sure and, second, I would hope if you have not visited ETS that you would spend at least 2 days there. You appear to be someone who wants to be very fairminded, so you should want to have an overview of the whole situation.

Two hundred seventy-five dollars is a complete waste. You are an intelligent young man. You can get this free and the books that David talked about for less than $20.

Mr. Geer, I do want to ask, you did not have the idea of throwing the baby out with the bath water, did you, when you were giving your testimony?

You cited a very serious case, and certainly something should be done about it when you look at the overall picture of the experience you have had.

Mr. **Geer**. My concern overall with testing is that you have an instrument like SAT, which is the one I am most familair with and which was designed to do certain things, has been extraordinarily carefully monitored and protected now over 35 years. I think the largest problem with the SAT is not the example itself, but the gross misuse of the SAT.

That test was never designed to evaluate American high school education, and yet the annual report of the decline in SAT scores is now being viewed as a measure of American high school education.

Mr. **Goodling**. Our legislation doesn't touch that.

Mr. **Greer**. As a person serving on councils, I protested as loudly as I could and told the College Board and ETS they had no right to use that information in that way, and they were using information that was on an examination, information that they had collected over 35 years, and they were using it in incorrect, improper, and manipulative ways, in my judgment.

Those are the issues that I am perfectly willing and very, very capable of arguing within the Councils of the College Board and ETS. I am also the president of the Illinois Advisory Committee on Non-Public Schools and represent 1,400 public schools, 370,000 children, and I don't want the Federal Government too involved in education.

I do not want a school board sitting in Washington. We have a very, very thoughtful and a very successful relationship between the nonpublic schools and the State school board in Illinois, and that is complicated enough for us in educating the school board and making sure they understand the rights of nonpublic schools and, therefore, the rights of the students in those schools in terms of testing.

Mr. GOODLING. I agree wholeheartedly with what you said.

Mr. GEER. It's time for a major revision of the SAT. You are perfectly right in saying the system that the SAT replaced in the later thirties, early forties did not serve American education and was clearly discriminatory.

The SAT was designed in the middle thirties and is an instrument that has served the education well, but is no longer, in my judgment, the kind of instrument that we should have for college admissions and for placement.

Mr. GOODLING. Does that take Federal legislation?

Mr. GEER. No; it certainly should be possible through the College Board and ETS and that instrument could be designed. One of the things that Dr. Howe talked about, as long as the SAT was a totally closed and secret instrument, students didn't even receive their scores until the middle sixties, there would be no way anybody would have access or understanding of SAT to even question it.

The testing and issues, the fact that much more information is now available, has been positive on the question of whether all the questions on all the SAT's should be published. I don't know. Every time a student takes a SAT, they contribute to the verification and experimentation with valid scores. I don't know how many thousands of valid questions ETS has. They talk about having 24, 25, 26 valid versions of the SAT. My guess is they also have literally thousands of other individual questions that are valid and I am, having served on the Council on Entrance Services, a little uneasy.

I think that is a somewhat definitive argument. I wish they would talk about whether that examination really meets the needs of American education. In my judgment, it does not and that it's an instrument, a good instrument an interesting instrument, but no longer as useful as it was in 1938, 1940, when it was designed.

Mr. GOODLING. I am sure you would agree with me; we tried to get teachers to use standardized tests as a tool to improve their teaching, to help young people with their shortcomings. Again I don't believe we need Federal legislation.

Chairman PERKINS. Mr. Weiss?

Mr. WEISS. Thank you, Mr. Chairman.

I happen to be one of those old-fashioned people who thinks that the educational process ultimately works. On the other hand, hearing some of the questions from this side of the table and some of the comments after the days and days of hearings that we have had on all sides of this issue, I tend to doubt my own judgments and conclusions.

Mr. Geer, you have been working with students. What is your opinion as to the effectiveness of coaching sessions?

Mr. GEER. Well, I alluded to my experience in Trenton which involved a kind of coaching, and I would agree that you have to define what coaching is. I think that experience with the SAT is extremely beneficial to a student, and you see this reflect in SAT scoring matrix.

They adjust the score downward every time a student takes a test, because the more familiar you become with that instrument the better your scores are going to be. You have to look at something like the SAT which is a modified intelligence test and, therefore, the more experience you have with that particular kind of test the more you will be able to respond to it.

Students do benefit considerably from experience with that test. Understanding the way they normally operate in school is not the way that you can operate on that test. If you have a student who reads a reading comprehension passage and is interested in understanding the material and wants to find the right answers, that student puts himself to an enormous disadvantage of that test.

You must develop the capability to move rapidly and quickly and accurately and efficiently from item to item, and not think about an item 5 minutes ago that you may have missed or you may be thinking about. There are special kinds of concentration, and there are special kinds of mental disciplines that are necessary for the SAT.

Unfortunately, they are not the kinds of mental disciplines that I as a person running a school would want my teachers to do in a normal classroom. They are contradictory to standard classroom proceedings.

Mr. WEISS. When the pamphlet that the Educational Testing Service sends out as being an indication that all you have to do is read this and that and that really takes the place of the coaching schools, you don't agree that that benefits us?

Mr. GEER. It certainly helps us. The fact that a student can take a test and have experience with the pacing and the timing that you have, there are 6 half-hour sessions on that test. One is a test of standard written English. One of them isn't even going to be scored, probably three match, or three verbal, and how a student paces himself, moves from section to section in that 3-hour period works under considerable pressure.

Mr. WEISS. Would you tell your students, for example, who are considering taking any of the established coaching schools, that it would be a waste of $275 if they took that?

Mr. GEER. No; I am very, very uneasy about an issue that Mr. Goodling did talk about. There are lots of unscrupulous people that prey on students and parents and citizens and present completely unprofessional and detrimental experiences in programs to prepare students for SAT's, and I am opposed to that. There is work that can be done, experience and practice, familiarity with the exam for a student to be able to know what his strengths and weaknesses are.

That is crucial that a student know when he can take a reasonable choice between two answers or when it is probably a waste of time. They are not going to do very well in that section, so it is best to cut their losses and move on. The other issue with the SAT is the fact that the student has to develop the skill in knowing when not to answer a question, because every question that you answer and get wrong reduces by a quarter, usually a quarter, reduces the number of questions that you have right, so a student must develop skills in knowing when it's reasonable, when they have a good chance of getting the question right and when they are doing themselves a disservice to answer the question.

We don't teach that in school. You don't get that in your daily classroom experience. My feeling working with minority students in Trenton, that is where they killed themselves. They were dutiful, hard working, thoughtful, earnest students. They would start on that SAT and get every single question right. They got about two-thirds of the test. They answered questions they have no business answering. They had a terrible time not answering any question, because they felt it was an insult to their intelligence.

How could they give me a test? I am a good student and work hard and if I don't answer this I am somehow saying something about myself. All of the qualities that made them superb high school students destroyed them on the SAT. They had to practice and go through the experience of what is that instrument like, how is it scored, and the students who did that, worked through that, their scores increased incredibly.

There were some students to whom that kind of practice never made any difference in their scores at all. You do need in my experience to have somebody that can lead you through. Some very bright students can pick up that practice book and you have the explanation of the questions, the scoring explanations in that SAT book which are materials that the College Board has never produced before and is extraordinarily useful.

It's a beginning, but I would like to see a lot more information like that. I would like to see the College Board put out a book with three or four SAT's in it with the kind of explanatory information they began to put in the book.

Mr. **Weiss**. Thank you very much.

First of all, let me commend you on the presentation that you made. Listening to you it was hard to acknowledge and recognize the fact that you are indeed a high school student. You were one of the best and most self-assured witnesses that we have had, not only in this hearing but in the hearings that I have sat through over the course of these past 3 years or so, and I want to thank you for your testimony as well.

I want to start off with a comment that Mr. Goodling made, if you want to take a trip to the ETS Center at Princeton, you might indeed find it interesting. I accompanied him on that trip, and I found a 300 or 400 acre, very modern center with a model, most modern imaginable, very nice. They have a mail order system, which makes a working politician's mouth water, because I would love to be able to have that kind of capacity.

Mr. **Drexeler**. Not only am I going to go, but our student government is going to sponsor a field trip to look at this.

Mr. **Weiss**. Don't expect that that is going to one way or the other tell you how the system works. It's a value judgment.

Let me ask you as to the coaching schools, how large is your senior class?

Mr. **Drexeler**. My senior class is about 350 students.

Mr. **Weiss**. What percentage of those students do you think have been or are taking any of the coaching courses that you referred to that you are taking?

Mr. **Drexeler**. I would say probably about one-third, but being from Great Neck, which is reputed as a very affluent community, to be able to take a course like Kaplan is much more easily afforded than in a community which is not as affluent, and that puts the students in Great Neck who can take it at a big advantage.

Mr. **Weiss**. Dr. Bond, in the course of the testimony today and in the course of testimony on this legislation throughout the past five or six hearings that we have held, I keep hearing comments about what is really required is improvement to the basic educational process.

I think you addressed yourself to that. I have no fault to find with that as being the major thrust, but focusing on improving the educational process, does that preclude taking other steps like reforming the process of the standard aptitude tests that are given?

Isn't the argument that we hear about really the problems, one of continuing education very much like the argument we used to hear before the Supreme Court decision in 1954, that separate but equal is OK?

That is not the problem. The problem is improving education.

Dr. **Bond**. Exactly, and the country is still desegregated. You are absolutely right that, yes, there are other things that can be done. My concern, however, is with the notion that item disclosure will somehow result in lessening, if you will, of the consistency on standard deviation, the differences that we find.

I predict that if there will be any change in this consistent difference, it will tend to increase any percent. My personal opinion is that those who will benefit from this particular section of the bill will typically not be black students.

I think the basic notion that, oh, I got this one wrong, and the real answer was No. 5. If it does not involve concomitant construction in the underlying principles of that right answer, that is kind of a superficial exercise.

I would like to comment briefly, if I may, on Mr. Geer's testimony, which I found rather enlightening. I have also had other instances at the University of Pittsburgh where that has occurred which tends to protect the rights of students.

I have seen this happen in the past, and I cannot be in any more agreement.

Regarding the deficiencies in multiple choice testing, I would agree that there are certain pressures and certain problems involved in multiple choice testing which are, for example, dismissed or obviated when you consider open-end questions, where the student actually has to provide the right answer.

It is this kind of thing that I am talking about when I say, OK, let's go to open-ended questions. I don't believe that will solve the problem either, at least technical problems will not solve the fundamental causal agent here.

I am currently involved in research which involves an alternative to multiple choice, dichotomously scored items. It essentially involves the idea that students should get partial credit for any information they have regarding a particular item.

I would be glad to send that research to the committee.

Mr. Weiss. Dr. Cahow, isn't your experience that earlier on when we have a situation like that which Mr. Ford referred to and there seemed to be so many more people applying for admission to colleges and universities, that he said at that time perhaps the tests were used to keep people out because they had to have some way of doing that?

Is it your experience that perhaps the tests were used that way 5 or 10 years ago?

Dr. Cahow. I cannot assume that that has not been the case in institutions where pressures were severe for admission, but I don't know of any cases, but perhaps there could have been an overemphasis on the scores.

Mr. Weiss. You commented in the course of your testimony on Dr. Keller's earlier testimony as to the norms used in relation to distinctions between male and female and scoring them separately, and that noting at the admitting schools as to what was happening.

Do you think that ought to have been done in relation to other areas of differences, be it race, ethnicity, or geographical differences, rural versus urban differences?

Dr. Cahow. Yes, I do.

Duke, along with the consortium of schools, have been asking for this information for several years because we think it is critical to our marketing routine to discover where these students are that we think can benefit from our school.

The question, who owns the data? ETA takes the position that the test score belongs to the student and is released to an institution upon permission by the student and may not be released to the outside world and ETS cannot release it to the outside world.

I have reason to believe because of my conversations with the college board and at ETS that they probably could release the data. They have been reticent to do so because of the incorrect interpretations that might be made because lower scores are achieved by one ethnic group than a nonethnic group.

Mr. Weiss. According to Dr. Keller's testimony, if I understood it correctly, what he was saying with regard to the GRE was that there is in fact a skewing done, there is an additional credit given to men because they are not expected to do as well on the basis of past experience on the verbal part of the test and that therefore in essence the men get an advantage because they have a certain position given to them within the test scores.

That seems not to have required anybody's approval. Why should it require anybody's approval to do anything in the other areas of differentiation?

Dr. Cahow. The men's scores are based on the men's norm and the women's scores based on the women's norm. So there is not any skewing to give men or women an advantage or disadvantage. You are merely comparing men's scores with all other men's scores so you can see where he ranks.

I don't believe it would require ETS to get any kind of special clearance to release this information in aggregate. I think it certainly would for an individual student's scores to be released.

The information that is released on the GRE is aggregate information. I don't think we are breaking any laws if ETS decided to do this in terms of linguistic background or economic background as Dr. Keller asked.

Mr. WEISS. Dr. Clawson, do you know how long the old MCAT tests were in effect? How long were the old MCAT tests used before the new ones were created?

Dr. CLAWSON. 16 years. They were not in effect when I went to medical school. They were put in subsequently, but they were in effect for many years.

Mr. WEISS. I assume that since the new ones have only been in effect since 1977, that really it is too early to tell as to whether in fact they will or will not prove out to be any, judged to be any better or any more valid by even the medical college community than the old MCAT?

Dr. CLAWSON. I think it depends entirely on whether the test is used appropriately or misused. Almost everything I have been hearing in the testimony this morning are misuses of tests. If it is misused and if it is thought that it will be the key predictor in the success of a person going into medicine, then it is going to not be a good test.

The only basic thing that went into it was a conceptual one in which it involved some 150 faculty and students. I might say the organization of student representatives worked and had great input into the construction of the new test. In the past the concept was that you tested in considerable depth in the subjects that were being tested.

The concept of the new MCAT is that you limit it to very specific types of courses, that is, the courses that would normally be taught in the first year of college and the questions were selected by student faculty panels as being the best questions.

It does not, though, aim to be the predictor, if you will, in terms of who is going to be a good physician. I think that is quite clear. That is not the point of it. It is to have an objective measurement to be able to better equate the grade point averages that students are presenting to you when you come in with the student's ability in those courses that we are testing in.

That is the point of it. The grade point average has got to be looked at as variable, depending upon the college that the person goes to. It is not an absolute term at all. This is, I think, the great advantage of the MCAT and the new formula.

Mr. WEISS. You correct me if I am wrong—there is not just one single new form, is there? There must be a series of new forms. It is not just one test of 150 items or questions. There must be a half dozen new forms of 150 questions, I would assume.

Dr. CLAWSON. There are far more questions than 150. The 150 were the ones worked on by the people putting the test together. There are far more tests than that, but not the exhaustive school. It is not large enough to prevent the students from memorizing the answers.

Mr. WEISS. Suppose you relate it to what happens to those tests and the experience that they had at his high school in the course of 1 year. Is there any reason to believe that in fact there will not be people sent in by the so-called cram schools who give courses on MCAT to memorize those questions, that there will not be instances where inadvertently or unintentionally some of those forms will end up reaching the outside world?

What happened in those instances, assuming that those are real possibilities?

Dr. **Clawson**. That is a possibility that concerns all of us, not only a possibility but a probability that that will happen. I think we have to leave it to the board of MCAT to be sure that they can over time adjust the questions enough over time so as to get some level of predictability as to the degree of what is happening.

I have great faith in that particular testing institution as being totally fair, trying very hard to look out for the student's interests as well as getting a standardized school.

I am sure that questions will be changed on an annual basis. But I think the testimony here of Mr. Bond is, to me, strikes home. It takes too long to put together a whole new test. That is why we want to be able to use questions over again, but we will use them in different patterns.

Mr. **Weiss**. Of course, you know that the testimony that we have received from the American College Testing Service, for example, which I think prepares the MCAT, stated that on their college admissions not the medical college admissions, but that they, I think four out of five sections on any particular test are composed of 100 percent brand new questions and, therefore, overall 80 percent of their questions on any particular test given are in fact new questions. So that I am not sure that in fact it is that kind of a burden.

But my question to you is: If in fact it is reasonable to pursue the kind of thing Kenneth was testifying to will occur with the medical college test as well as the others, aren't you in fact giving an advantage automatically to those students who can afford to go to the cram schools to the disadvantage of those students who cannot afford to go to the cram schools?

Dr. **Clawson**. You are bringing up an issue that is really rather foreign to me where I happen to be working.

Well, different courses, if you want to call them cram courses, or that, to any student who comes to us who wants to spend the time and efforts.

We have faculty that we hire to work with groups that feel for whatever reason they may be underprivileged. We do give courses for any student that wants to come to us on test taking on how to take standardized tests. It is an open thing. It is not an expensive thing.

So in my area of the country I do not hear this as being a great problem. We have an open door policy and we will help students who want to prepare themselves to take the MCAT tests. We put on night courses and do everything to open it.

So this is something that I must say I had heard about, heard about in particular sections of the country. It is not a big problem for us. The answer is clearly, if the school that a student is applying to puts tremendous weight upon the MCAT test, which ours does not unless there are other factors that make us want to look at it in great depth, then students are going to try to take any advantage they can of anything in order to make their MCAT's better.

So I think you have set something in motion and I don't think that the bill and the release of the questions as such is going to really be of that much benefit to the students over the opportunity to work with faculty to help them develop the skills necessary for both taking the test and answering the questions.

Mr. **Weiss**. Mr. Gastfriend, did you want to comment?

Mr. **Gastfriend**. Yes.

First of all, I think again it is to Dr. Clawson's credit and to the credit of his school that he can say these kinds of things about his medical college.

I think that by and large that is in the minority of medical school setups. I want to point out that what I was quoting about our position about the MCAT I had taken from an AAMC press release quoting from John A. D. Cooper: "The tests are essential. They are designed to predict how well a student will perform in medical school and later as a practicing physician."

Mr. **Weiss**. Indeed that press release was sent to us on the first day of the hearing.

Mr. **Gastfriend**. OK.

Another thing I want to point out as to preparation, the national board exams which are given to medical students and postgraduate training physicians issue a series of tests in subject areas.

They also issue what some schools call a miniboard prior to the tests which are given after 2 years of medical school. These tests are very commonly used for the purpose of preparing the medical school's sophomore class to take the boards. They are given 4 weeks in advance and scored quickly and given back to the students and they use this to practice the testing situation.

So I think medical schools at least are well aware of the need of preparation and learning the system of test taking.

The last thing I should mention, I am sorry Mr. Goodling is not here, being from Pennsylvania, but last week in the Pennsylvania State Senate a bill similar to the New York one I am told was introduced with 10 out of 18 members of the education committee cosponsoring the bill.

Mr. **Weiss**. Thank you, Mr. Chairman.

Chairman **Perkins**. Mr. Andrews?

Mr. **Andrews**. Thank you, Mr. Chairman.

Obviously, the hour is quite late and the subject very complex. I would just comment, first of all, that I have been tremendously impressed and confused by the opinions that the witnesses we have had today and numbers of those we have had previously exhibit in my estimation, both tremendous comprehension of the subject and in-depth commitment to hold it as it is and the best interests of what they consider to be the right perspective or to change it for the same reason.

And to try to, as a member of the committee, look upon this in a rather judicial way, having a great deal of respect for the people's ability and their motivations, and then to have to decide on one side or the other when they differ so extensively is quite difficult for me, knowing less about it probably than any witness here.

But I can't help but conclude that, as is the case with most legislation, that apparently whether the given witness concludes that this is a good bill or whether it is not doesn't seem therefore to rest upon the information the various witnesses have or their commitments to what is right, but rather the perspective they have as to what the mission of the school is.

Now that is the only conclusion I can come to. It seems that each has a perspective, depending upon where that person comes from. One group seems to see the mission of the school as having a social benefit. But what social benefit can the school give to society as a whole?

Other witnesses seem to think that the primary mission should be the civil rights of applicants for admission to a given school.

Others don't seem to think it is a matter necessarily of civil rights in terms of necessarily who can score the best and by the fairness measurement, but rather if they are minorities or women or not, et cetera.

But I can't but go back to my own experience just as I think most witnesses go back to theirs. The only mention that has been made just in passing that relates to my own was made by Dr. Clawson when she, I think, just parenthetically said that a certain number of years ago it was determined that in eastern Tennessee more doctors were needed and that is why the medical school was established in the first place.

In my experience which with this subject is rather brief, but as a member of the North Carolina Legislature, I can remember so well the fight as to whether there should be a medical school added at east Carolina because doctors were needed in eastern North Carolina, the more rural areas rather than in the Piedmont. Hence, you should geographically have a school in the east.

I can well remember the fight to double the entrance to the medical school at Chapel Hill because, as you say, it was conceived that more doctors were needed generally back then which was in the sixties.

So I can't but conceive that the schools basically—I think that all perspectives here are valid and they should all be considered. The schools are publicly supported. Individual members of the public, be they students or parents or whomever, have a legitimate, I think, right—and I agree with that.

The total distribution can't just be the school or the need for doctors in a given area. But I still think that has to be paramount. I think that is really what the schools are paid for primarily by the public to accomplish and that is to afford doctors or lawyers or whatever, the social workers or whatever is sought to be accomplished ultimately, or people who are trained, not necessarily the civil rights of those to be trained.

I don't say that is not important. I think it is a legitimate concern. But I don't think that is the end unto itself. That is not to say it should not be given due and fair consideration, but I believe the real objective ultimately, not the only one but the principal one, is that the schools should produce for the community not necessarily social improvement or civil rights but instead the mission of the school I think primarily is to produce people who in turn can produce whatever goods or services are desired and hence that is why the school was established.

From that point I have to be tremendously influenced by the fact that the schools, then, I would think, should themselves determine whether tests should be given and under what circumstances and whether they should be made public and to what extent, if any, they should be made a criteria in the determination as to whether a given applicant should be enrolled.

If that is the case, I don't see how that the Congress as representatives of a total public can fairly or adequately—and I as a member give due and equal consideration to student rights, civil rights, social change in terms of minorities, and all the other many things that we as Members of Congress are supposed to be accountable for.

I think instead we have to consider that we are basically dealing with schools and it just seems to me that, therefore, the resolution of this problem has to primarily rest, therefore, with the schools rather than to be usurped by we, the representatives of the total public.

That is about as good a way as I can describe the situation in a short time, felt some obligation to attempt to enunciate the way I felt about it.

Thank you.

Chairman **Perkins**. Let me thank all of you distinguished witnesses for your appearance here today. I know that this is very controversial and it is receiving considerable thought from the committee and it will continue to receive considerable thought.

But you have been most helpful and I hope that we can come to the right conclusion when we do commence to mark up the legislation. You have been most helpful. I appreciate your coming.

Mr. Weiss?

Mr. **Weiss**. Thank you, Mr. Chairman.

I would simply like to ask unanimous consent to have entered in the record an article that was referred to in an earlier hearing. This is by Dr. Michael Wallach of Duke University.

Chairman **Perkins**. Without objection the article will be entered into the record.

[The information referred to above follows:]

Jan.-Feb. 1976
p. 57-63

76001767

Michael A. Wallach

(7)

Tests Tell Us Little about Talent

CRS-MASTER FILE

"*Although measures of academic skills are widely used to determine access to contested educational opportunities, especially in their upper ranges they lack utility for predicting professional achievement*"

Quality educational resources are an inevitably scarce commodity, and the decisions as to which young men and women will gain admission to highly regarded colleges and universities and receive scholarships or fellowships while they are there are difficult ones. How should these decisions be made? The relative merit of candidates is the usual criterion, but the definition of merit is problematic. Typically, "merit" is defined as scores on tests of intellectual abilities and grades in the customary types of academic course work, which themselves are usually determined by some kind of test.

Of course, test scores and grades are not indices of merit in their own right; they are thought to provide a shorthand indication of a student's competencies in the world outside testing and classroom situations. Recent research on the nature of talent indicates, however, that the premise that tests reflect the potential for achievement is false for scores across the upper part of the range on customary assessments of academic skills—precisely that part of their range in which such scores are most often used for selecting recipients of the most contested educational opportunities. The answer is not, as I and some others first thought, to replace tests aimed at assessing academic skills with tests that would reflect a person's "creativity"; rather, tests should be used only to screen out candidates who score too low. To make distinctions among the candidates who remain, we should rely not on tests but on samples of professional competencies themselves.

Disenchantment with the utility of tests in their upper ranges for predicting the sought-for competencies is not limited to the study of talent but has been increasing among most psychologists who attempt to devise ways of assessing human behavior. Traditionally, after theorizing about the relatively abstract dispositions that presumably underlie the display of the behavior, the psychologist attempts to develop tests that would signal the operation of these dispositions. Since the connection from test to criterion is mediated by complex theoretical constructions, the former need have nothing in common behaviorally with the latter. A psychologist could try to evaluate the chances that a particular therapeutic treatment would succeed with a subject by something as unrelated as the way in which the subject completed sentence stems or interpreted ambiguous pictures.

With enough negative evidence in hand from attempts of that sort, we now know (see e.g. Mischel 1972) that more reliable answers are provided by assessing what the subject does in a sample of the treatment situation itself. The problem is that the test responses, even if they possess a modicum of "criterion validity"—that is, give results better than chance at predicting the behavior of direct interest—inevitably reflect other factors as well (see Wallach 1971a; Wallach and Leggett 1972). And the greater the conceptual distance between the test and the performance to be predicted, the less reason there is to believe that the test will tell you what you really want to know. It will tell you about the person's response tendencies in situations that resemble the test rather than in situations that resemble the criterion.

Academic skills tests

Above intermediate score levels, academic skills assessments are found to show so little criterion validity as to be questionable bases on which to make consequential decisions about students' futures. What the academic tests *do* predict are the results a person will obtain on other tests of the same kind. Some of the evidence for this statement is based on accomplishments of directly meritorious kinds during the student years, and we shall consider it first. After that, we will turn to accomplishments in occupational and professional roles. The research along both lines is extensive enough that we can only sample it here—but the sample is representative. Further amplification of this evidence and its meaning can be found in Wallach 1971a, 1971b, and in press; Wing and Wallach 1971; and Wallach and Wing 1969.

Most of what constitutes the usual academic course work, with its di-

Michael A. Wallach is Professor of Psychology at Duke University. He has also held faculty positions at Harvard, MIT, and the University of Chicago. He did his undergraduate work at Swarthmore and is a Harvard Ph.D. Dr. Wallach, who served as editor of the Journal of Personality from 1963-72, is a past president of the Division on Psychology and the Arts of the American Psychological Association. His research has included the study of cognitive competencies, talents, and skills in children and adults, with particular reference to implications for improving educational policies and practices. He is coauthor of a work on compensatory education, Teaching All Children to Read, *which will be published in early 1976 by the University of Chicago Press. Address: Department of Psychology, Duke University, Durham, NC 27706.*

gesting of a more-or-less formal curriculum which is then checked by memory-oriented examination questions, differs in kind from most vocational roles. The cramming and test-taking skills needed to obtain good grades in such courses, however, have much in common with the skills that should help earn high scores on tests of intellectual ability, and it should not surprise us to find scores on "intelligence" tests correlating rather well with "academic achievement" indices—a point that has been thoroughly documented (Tyler 1965; Wallach and Kogan 1965; Wing and Wallach 1971). By the later high school years, however, students are capable of producing work of intrinsic significance in a vocational sense. Sometimes—though not always—this work is accomplished outside the school curriculum.

In a study by Parloff, Datta, Kleman, and Handlon (1968) the novelty and effectiveness of research projects conducted by high school students independent of course work were evaluated. Differences in quality among the projects were substantial—but were unrelated to intellectual aptitude test scores or to high school grades. Even course grades in *graduate* school often seem unrelated to the competencies associated with professional performance. Using a judgment scale taken from Taylor's (1958) work on the quality of research by professional scientists, Mednick (1963) asked faculty members to evaluate research projects conducted independently by psychology graduate students for the imaginativeness of the project and the degree to which it constituted a contribution to knowledge—reasonable criteria of competence for professional work. The judged quality of the projects ranged widely, but differences in quality had nothing to do with the grades received by the students in their courses or with their scores on the Miller Analogies Test. Scores on this test, however, predicted the course grades.

These results typify the findings of a wide array of studies of high school, college, and graduate-level students concerned with assessing accomplishments that anticipate post-schooling achievements. Public recognition by knowledgeable evaluators is a major criterion of quality in this research—and, while admittedly imperfect, it is reasonably close to the way in which professional excellence is judged in life. Attainments in science, literature, art, music, and dramatics have all been studied, and indications of quality include, for example, publication of a scientific paper in a professional journal or of original writing in a nationally circulated periodical or a book. While sometimes direct evidence of such attainments can be obtained (e.g. Morrison 1963, where direct judgments of competence in dramatics are made), it also has been found that student reports on these matters are sufficiently accurate to provide a usable source of data. Maxey and Ormsby (1971), for example, demonstrate high agreement between student reports on out-of-school accomplishments and information provided by the school—even when college admission is at issue for the students.

The results of these studies show that, for instance, for the more than 500 undergraduates studied by Wallach and Wing (1969), accomplishments outside the classroom in each of the seven fields explored were about as frequent among students in the lower third as among those in the upper third on the College Entrance Examination Board's Scholastic Aptitude Test (SAT). The fields chosen were intended to span a representative array of the accomplishments beyond the classroom for which school is expected to serve as a preparation. Some of these activities—work in literature and science—are close to traditional academic disciplines, while others, like music and art, though not traditionally academic fields, are still within the customary purview of most higher educational institutions. Still others—dramatics, for instance—may or may not be included in a college's curricular spectrum. For none of the areas studied—literature, science, art, music, dramatics, political leadership, and social service activities—was accomplishment related to SAT scores, either considering the sample of students as a whole or taking men and women separately. Furthermore, the extent to which students exhibited achievements in more than one of the seven areas was likewise unrelated to SAT scores. These findings hardly support the idea that the SAT in this upper range reflects brightness broadly defined. On the other hand, students with the higher SAT scores were found to have higher course grades.

Even more compelling data pointing in the same direction come from Wing and Wallach's (1971) studies of college applicants. In that work, the subject pool was broadened to include not just undergraduates but high school students who had applied for admission. Definitions of accomplishment outside the classroom were restricted to relatively rare achievements of merit—for example, publication of original writing in a regularly circulated nonschool publication or the receipt of an award for such writing; having won top prizes in state or regional science competitions or having been chosen to attend a summer science program sponsored by the National Science Foundation; having won a prize in a musical competition or having done professional work in performance or composition.

The SAT scores of the more than 4,000 applicants of both sexes studied in our research on this question tell us very little about who is more likely to excel at accomplishments of these kinds. We found that if candidates are selected from the applicant population according to a hypothetical decision rule that favors applicants who scored high on SAT tests, applicants with the sorts of accomplishments just noted receive little or no differential benefit. Yet the selections actually made through the admissions process were found to be predominantly a function of the SAT score results. Similar results have been found in a study by Holland and Richards (1965), among others.

An implication of the foregoing is that score distinctions within the upper range should fail to predict the people who have demonstrated meaningful achievements in occupational and professional roles after leaving school, and this too seems to be the case. This outcome is so uncomfortable that evidence along these lines can meet with resistance. Take, for example, a careful study by Harmon (1963), in which the professional contributions of some physicists and biologists were evaluated by scientist judges on the basis of such criteria as publications and patents granted. The evaluations, by three or more experts working independently, were about as sophisticated and meaningful as such judgments could be. They were compared with college-level academic proficiency data—customary kinds of verbal and mathematical aptitude test scores, the score earned on an advanced achievement test in the student's field of concentration, and grades obtained in science courses. How good a professional scientist the person became could not be predicted from any of this academic proficiency information—and indeed, for nearly half the correlations computed, the direction of the relationship was negative rather than positive. Yet Harmon's reaction to this finding is to question not the predictors but the criteria!

The same result has emerged in study after study, whether intelligence test scores or course grades are used as predictors. Helson and Crutchfield (1970) found for mathematicians, for example, that the IQ scores of those chosen by knowledgeable peers as doing particularly good research work were no different from those of controls. Chemists and mathematicians chosen by peers for the quality of their contributions were found by Bloom (1963) to score no higher than controls on a variety of intellectual ability tests. Further data on research scientists from MacKinnon (1968) yielded the same outcome, and comparable results emerge from studies of psychologists (Marsten 1971), artists (Barron 1963), and architects (MacKinnon 1968), as well as from a range of evidence reviewed by Hoyt (1965, 1966) on occupational activities in scientific research, engineering, medicine, and business. In Great Britain, Hudson (1966) similarly finds undergraduate academic distinctions having little relationship to subsequent occupational eminence among scientists.

Although critics like Weitzman (1972) seek to question such findings on the ground that these studies consider too narrow a range of academic skills differences, the range is broad enough to include many more individuals who *seek* superior opportunities for educational advancement than can be granted them, and the question of how, within this range, selections are to be made is a real one. The lack of predictability seems all the more striking when we consider that a self-fulfilling prophecy effect no doubt exists in some degree, leading students with superior test scores and grades to be preferred for superior occupational slots over those with less impressive academic credentials. The evidence would seem to be all the stronger, since it must be emerging in spite of a prophecy bias of that kind.

"Creativity" tests

Extensive attempts to assess "creativity" or "originality" gained momentum in the 1960s, leading quickly to the need to make distinctions between tests that were simply assessing for intelligence once again but under a more fashionable label and tests that correlated minimally with conventional tests of intellectual skills (see Wallach and Kogan 1965; Wallach 1968). The tests that seemed best to meet this requirement (Wallach 1970 provides an extensive sifting of the evidence on this question) focus on the ability to produce in response to a given task a large number of ideas that are reasonably appropriate and, at least partly because of their number, include a rather large proportion of relatively unusual responses. Surprisingly, "ideational fluency" in this sense showed, for those above an intermediate level on academic skills, little correlation with standing on intelligence tests.

The ideational fluency tests typically asked the respondent to generate possible ways in which specified pairs of objects are similar (such as a cat and a mouse), or possible uses for some ordinary object (such as a newspaper), or possible instances of a specified category (such as things that move on wheels), or possible interpretations for various presented abstract line forms (such as a triangle surrounded by three circles). Time and again it has been demonstrated that samples ranging from intermediate to high on customary intelligence indices show little or no relation to the number or unusualness of the ideational responses on tests of this kind (see e.g. Wallach and Kogan 1965; Wallach and Wing 1969; Wallach 1970, 1971a). More unusual ideas tend to come later in the sequences of ideas, so that unusualness seems to ride the back, as it were, of the tendency to produce longer ideational strings, and respondents are relatively consistent in their degree of ideational fluency across different kinds of tasks.

Explanations for the observed differences in ideational fluency included varying tendencies to entertain the hypothetical (Wallach 1967) or to pick up information from the periphery of one's attentional field (Wallach 1970). Introspective accounts by highly talented writers, scientists, mathematicians, musicians, and artists of the importance for their creative achievements of letting ideas flow abundantly in an associative manner are also in evidence. They were far from decisive, of course, since ideational fluency could also characterize less talented contributors to the arts and sciences. All of this did serve to raise hopes that ideational fluency tests would be more helpful than academic skills tests for making prognoses about future accomplishments.

Some limited evidence can be found for the success of these tests. In Wallach and Wing's (1969) college student sample, those scoring in the highest third on ideational fluency showed more out-of-school attainments in literature and science than those scoring in the bottom third. At best, the linkages were tenuous, however, and for some kinds of talented behavior no links to ideational fluency were found at all. Another shred of positive evidence is Singer and Whiton's (1971) finding of a bit of linkage between children's ideational fluency test scores—but not their intelligence test scores—and the judged expressiveness of their drawings. In a related study, Wallbrown and Huelsman (1975) found comparable positive evidence for a clay product but not for two crayon drawings made by children. Attempts to provide criterion validation for ideational fluency tests by Cropley (1972; see also the reanalysis of Cropley's data by Jordan, 1975) and by Kogan and Pankove (1974) seem quite negative in their outcomes.

Ideational fluency tests may be minimally related to intelligence tests, but they are little if any better at predicting the achievements we care about. Subjects vary widely and systematically in their attainments—yet little if any of that systematic variation is captured by individual differences on ideational fluency tests (although the tests too show wide and consistent differences from person to person – well beyond whatever small portion of those differences may relate to significant attainments). Again, we are facing the multiply determined nature of the test responses. But the inferential gap between test responses and writing or science attainments, say, is huge. It is too easy for other factors unrelated to the attainments we want to predict to influence the test responses—perhaps overwhelmingly.

Generating larger numbers of task-appropriate responses, and reaching thereby more that are statistically infrequent, can simply reflect greater conformity to the implicit or even explicit suggestions of an authority figure—the tester—that this kind of behavior be provided. It also can reflect obsessive hair-splitting—the making of finer-grained distinctions among responses that someone else is aware of but simply ignores, as, for example, in enumerating the kinds of things a coat hanger could be used for. Thus, Bowers and van der Meulen (1970) found with college students that individual differences in ideational fluency test scores correlated positively with scores on a scale presumed to measure susceptibility to hypnosis. The finding that those who seem more suggestible in response to authority figures—the susceptibility to hypnosis measure—also earn higher scores on ideational fluency tests seems to indicate that those higher in suggestibility try harder to provide what the tester seems to want on the ideational fluency tests. Yet the authors, taking ideational fluency scores as a functional equivalent for creative attainments, interpret the data to mean that creative people are more suggestible!

If academic skills tests in their upper ranges tell little about vocational competencies, "creativity" tests in this middle- to high-intelligence score range are not about to fill the gap. Even if creativity tests *had* proved to possess the hoped-for criterion validity, making them sufficiently impervious to direct coaching to be useful for practical purposes would have been difficult if not impossible.

From tests to attainments

We recognize talent from the display of excellence at vocational pursuits, and it is these significant accomplishments as such that offer the most legitimate basis—once certain academic skills qualifications have been met—for the judgments of merit that must accompany the allocation of educational resources. But how could this approach be implemented? My recommendation is not to abandon traditional ability tests and grades, but to limit their role to the screening out of those who score too low. In Wing and Wallach's (1971) use of direct talent assessments in college admissions, for instance, we apply cut-offs that reject candidates from further consideration if their SAT scores place them in the lower half of the distribution for high school seniors who go on to college or if their grades place them in the lowest third of their class. Choice of these particular cut-offs was based on projections indicating that those above the cut-offs should be able to do the academic work expected of students at a highly selective college. For those above the cut-offs, SAT scores and grades receive no further attention in making preferential selections for admission or scholarship support, for, as we have seen, it is within such upper ranges that test score and course grade distinctions seem to be without validity for predicting accomplishment. There is nothing absolute about the particular cut-off points; but some such thresholds should be specified and further increments above them ignored.

To select among students who fall above the academic cut-offs, we must specify the kinds of life achievements deemed meritorious in their own right; the ones chosen may vary depending on the purposes of the selection. It will be easier to select for graduate and professional schools than for college admissions, since a more definite target of professional competence is present. For undergraduates, the best approach is to take as evidence of merit achievements in whatever pursuits a college says it prepares its students to follow, restrict consideration to achievements of high quality, and give priority to candidates with noteworthy achievements in more than one realm.

Two classes of evidence should be gathered: verifiable signs of public recognition provided by qualified judges and work samples to be evaluated by the selection committee itself. Candidates in literature would be asked to provide verified information on work they have had published in nonschool publications, on the nature of those publications, so that the kind of appraisal that went into the publication decision can be assessed, on awards they have received for written work, and on the nature of the awarding agencies, so that again an appraisal can be made of the worth of the recognition. Candidates would also be asked to provide a sample of their best written work, to be judged by appropriate faculty members of the selection committee.

For prospective scientists, signs of public recognition would include verified information on research published in scientific journals or other nonschool publications, on the nature of these publications, on research awards received in science competitions, and on the nature of the awarding agencies. Candidates would also be requested to submit reports on scientific projects they have conducted, to be evaluated by suitable selection committee faculty members. The achievements may or may not have their point of origin within a school's curriculum; for college applicants the work that qualifies is more likely to be extracurricular than for applicants for graduate school.

How are candidates to be apportioned among different fields? At the graduate level, the answer depends on each department's goals and resources. At the undergraduate level, one feasible plan would be as follows (Wing and Wallach 1971): Define all the fields an institution wants to represent in its training. List applicants (above the academic skills cut-off) in descending order of work quality within each field, but give priority in the listing sequence to applicants doing high-quality work in more than one field. Then proceed around the fields in carousel fashion, taking the person most qualified from the first field's list, then from the next field's list, and so on, removing a candidate from any other lists in which he or she may appear as soon as the candidate has been chosen from one list. After choosing the most qualified person in each field, select the next most qualified person in each field, and continue until the available slots have been filled.

An institution can define fields as broadly or as narrowly as it wishes. It may have only two carousel positions—say, the sciences and the humanities; or it may have positions for criticism in literature and the arts, historical essays, creative writing, physics, biology, chemistry, psychology, and sociology. It may include performance in art and music or restrict itself to history and criticism in these fields. It may weight the fields in any proportion it finds appropriate, taking two people who qualify in literature and the arts, say, for every one in the sciences, or vice-versa, basing the decision on its feelings about its strengths and mandate as an institution, what higher education should concern, or what the society needs. But it must make its choices clear both to itself and to its potential applicants, so that the forms of competence it undertakes to select for and nourish will be reasonably explicit and hence available for examination and critique.

The objection might be made that these proposals would be time-consuming to put into effect. But selection committees already spend large amounts of time and effort in reaching their decisions. The difference is that most of their present procedures are relatively unsystematic, and the apparent objectivity of test scores can influence them more than it should—and more than they believe it does.

It might also be objected that some students will try to develop certain competencies in order to gain admission or scholarship support at a more selective college or university. But this criticism brings home the importance of concentrating assessment efforts on inherently meritorious activities rather than on measures that take their justification from the hope that they can predict these activities. Students will always feel some pressure to do the things that will increase their educational chances. Our obligation is to make the sources of such pressure as valid as possible—to reward students for meaningful displays of competence rather than for excelling on meaningless tests.

Besides their intrinsic significance, do the accomplishments we have considered also point to what people will do in the future? Richards, Holland, and Lutz (1967) give a clearly affirmative answer to this question across the years from high school to college. In a study of nearly 4,000 high school students of both sexes who were followed up after one or two years at diverse colleges, data on out-of-school achievements in literature, science, art, and music were collected for the high school and college periods, along with data on high school and college grades and scores on conventional academic skills tests taken at the end of high school. Once again, high school grades and test scores tend to intercorrelate, and both tend to predict college grades. Neither test scores nor grades, on the other hand, predict which students have excelled in nonclassroom projects in the sciences and the humanities in high school or in college, but achievements in some field while in high school tend to give a good basis for predicting that the person will continue to manifest accomplishment in that field while in college.

A recent study by Munday and Davis (1974) gives results comparable to those of Richards et al., but this time comparisons were made across a longer time span—from the end of high school to six years after that, typically two years after college graduation. Academic skills test scores, high school grades, and college grades relate to one another but do not relate to out-of-school achievements during the high school years and do not predict such achievements six years later. But out-of-school achievement in a given field during high school predicts achievement in the same field six years hence. It does so, furthermore, to about the same degree as academic measures at the end of high school predict college grades—even though in the former case we are talking about a six-year span and in the latter a four-year span. And the available evidence (see MacKinnon 1968) suggests that accomplishments outside the classroom during the student years predict who will be making substantial contributions in professional work later on—the information this time coming from what adult professionals say about their past behavior. Evidence of high-quality professional contributions at an early point in one's career, in turn, gives a good basis for predicting them later as well (see Albert 1975)—so from high school on, competence seems to beget competence.

Would it make a difference?

The catalogues of the selective colleges claim to seek students who have expressed their talent outside the classroom in the ways we have been considering. To determine to what extent actual practice matches this ideology, Wing and Wallach (1971) conducted studies of actual admissions decisions which included consideration of highly selective colleges—institutions where, if anywhere, out-of-school achievements could in principle play a role in selection, since these colleges are heavily overapplied and most applicants are reasonably well qualified in terms of academic skills. The results indicated that even at these colleges selection decisions tend to be monolithically oriented toward conventional academic standards. Knowledge of as simplistic an index as an applicant's score on the verbal aptitude section of the SAT lets one predict rather accurately the likelihood of the candidate's being accepted or rejected at both selective and unselective institutions. The difference between the two kinds of colleges seems not so much whether they depend on SAT-verbal scores in making admissions decisions, but on the level of score required for acceptance.

For instance, Wing and Wallach (1971) found that, at a random sample of 45 colleges, probability of acceptance increases more than fourfold when comparing applicants with SAT-verbal scores falling within score intervals of 350-399 and 550-599 (scores can range from 200 to 800); while at a sample of 13 highly selective colleges, the same order of increase in likelihood of acceptance is found when comparing applicants whose SAT-verbal scores fall within score intervals of 500-549 and 700-749. SAT-verbal scores in the 550-599 interval put a candidate at approximately the 80th percentile of high school seniors who go on to college. Yet likelihood of acceptance at the highly selective colleges sampled is *three times as great* if the applicant's score falls in the 700-749 interval than in this already high interval of 550-599. This means that, while relatively unselective colleges use SAT-verbal scores as a basis for screening out candidates whose scores are too low for them to cope with college work, the highly selective colleges use them to discriminate among applicants who already cluster around the high end of the measure. Marginal increments in SAT-verbal scores within this upper part of the distribution determine acceptance versus rejection.

Indeed, Wing and Wallach further discovered that predicting admissions actions at a typical highly selective college in terms of combined consideration on an equal weighting basis of the candidate's verbal and mathematical scores on the SAT along with his academic achievement level in high school led to such a high degree of accuracy in accounting for acceptances and rejections as to leave little room even in principle for other factors to play a role. About 80% of all decisions, and about 80% of all acceptances as well, were correctly predicted this way—even though the possible predictive accuracy in these situations is lower than 100% because the decisions to be predicted are themselves made with less than perfect reliability. This general result was replicated across two successive recent years' applicants for admission to the freshman class, with more than 4,200 applicants one year and more than 4,600 the next, in spite of the fact that one year's applicants were required to provide more extensive information on out-of-school accomplishments.

Admissions practices seem to be substantially at variance, then, with the diversity of talents that highly selective institutions say they seek. Could selection be carried out instead in a way that would genuinely implement this goal? Wing and Wallach explored this question by applying a hypothetical decision rule that gave preference to candidates who had demonstrated substantial accomplishments in the fields outlined earlier. The rule also gave preference to candidates as a function of the number of fields in which such accomplishments had been demonstrated. Finally, it permitted candidates to be accepted this way only if their SAT scores and grades were high enough (SAT scores in the top half nationally for high school seniors who go on to college, and grades in the top two-thirds of the high school class). If one selects by these criteria, the hypothetical accepted class overlaps only 40-50% with the class actually accepted. (Remember, by contrast, that selecting for candidates with the highest SAT scores and grades in high school results in an accepted class that overlaps about 80% with the class actually accepted.)

What are the applicants like who are accepted by the rule that emphasizes professional accomplishments? In the Wing and Wallach results, out-of-school science achievements of high quality are found for 15.6% of the accepted males compared to 8.4% of the male applicants, and for 15.5% of the accepted females compared to 7.4% of the female applicants. Out-of-school writing achievements of high quality are found for 19.3% of the accepted males compared to 9.5% of the male applicants, and for 33.8% of the accepted females compared to 14.6% of the female applicants. The accepted group is comparably accomplished in the other fields of activity that were assessed; these achievements are represented with far higher incidences than their rates of occurrence in the applicant pool. Basing selections on conventional academic criteria, on the other hand, or considering the selections made by the actual admissions committee, yields an accepted group far less distinguished in its achievements.

One defense of the use of standardized tests like the SAT for admissions purposes is that they are nondiscriminatory as such, and perhaps less discriminatory than any alternative objective basis for selection would be. The Wing and Wallach data give us an opportunity to compare for an applicant pool the effects on nonwhite candidates of using a hypothetical decision rule based on SAT scores and one based on accomplishments once certain minimum criteria as to SAT scores and high school grades are satisfied. Use of the SAT decision rule was found to be drastically discriminatory in its consequences: almost all nonwhite candidates in the applicant pool would be excluded from the group accepted. Use of the rule based on accomplishments, by contrast, was found to lead to a quite different outcome: nonwhite candidates would be accepted in about the same proportion as their incidence in the applicant population. Thus, using accomplishments rather than test scores as the basis for selection seems, if anything, to be less discriminatory in its consequences for nonwhites.

I believe that our proposal would produce a college class qualitatively different from those selected by conventional means. Does the difference hold implications for the society? Darwin argued that biological characteristics prosper across generations to the degree that they are represented in the organisms that breed. Something analogous may go on in educational selection: If conventional test-taking skills more than significant competencies govern access to contested educational opportunities, these competencies will be underrepresented in the people who win more valuable educational credentials in the society. Having the credentials tends to open doors to roles of greater influence, so the attributes of the people holding such roles will be more likely to influence the overall direction of the society.

Perhaps the reason why the more selective colleges and graduate schools claim to seek students who demonstrate diverse talents, while in reality choosing the reverse, is the extent to which academic skills tests and course examinations are rather automatically construed by faculty and administrators as reflecting a person's "intelligence" or "brightness" in all the senses of the term. Obtaining the brightest students possible becomes a way of defining the quality of an institution, and when colleges and universities judge one another's national standing in terms of the average SAT scores of the incoming freshman class or the average GRE aptitude test scores of the incoming graduate students, the heavy use of such yardsticks in admissions should be no surprise. As Wing and Wallach (1971) and Campbell (1971) have pointed out, furthermore, the faculty members at such institutions have themselves been selected for educational advancement largely by the very criteria whose use we have questioned. It is natural for them to perpetuate and see virtue in the ways of defining merit that led to their own rise—even when faced with contrary evidence.

The role played by testing agencies in preserving the established ways of doing things must also be mentioned. They have, after all, a vested interest in the continued and even amplified use of the testing devices they know how to make. To justify continued emphasis on tests like the SAT and the GRE on the grounds that they predict course grades, when neither the test score nor the course grade differences in their upper ranges predict actual accomplishment, is questionable to say the least.

The irony is compounded when educational commentators like Jencks and Riesman (1968) urge even heavier use of academic aptitude test scores instead of grades in selection on the ground that the former are less subject than the latter to irrelevant sources of bias such as teachers giving higher grades to students who are more polite. For it is their correlation with grades that provides the aptitude tests with their basic justification in the first place. But to deal with real accomplishments, once certain academic ability thresholds have been crossed, is to restrict the use of tests to a much more circumscribed function than testing agencies envision. Testing agencies should perhaps devote less effort to tests and more to helping educators assess achievement in activities that we value.

References

Albert, R. S. 1975. Toward a behavioral definition of genius. *Am. Psychologist* 30:140-51.

Barron, F. 1963. *Creativity and Psychological Health.* Princeton: Van Nostrand.

Bloom, B. S. 1963. Report on creativity research by the examiner's office of the University of Chicago. In *Scientific Creativity: Its Recognition and Development*, ed. C. W. Taylor and F. Barron. NY: Wiley. Pp. 251-64.

Bowers, K. S., and S. J. van der Meulen. 1970. Effect of hypnotic susceptibility on creativity test performance. *J. Personality and Social Psychology* 14:247-56.

Campbell, D. P. 1971. Admissions policies: Side effects and their implications. *Am. Psychologist* 26:636-47.

Cropley, A. J. 1972. A five-year longitudinal study of the validity of creativity tests. *Developmental Psychology* 6:119-24.

Harmon, L. R. 1963. The development of a criterion of scientific competence. In *Scientific Creativity: Its Recognition and Development*, ed. C. W. Taylor and F. Barron. NY: Wiley. Pp. 44-52.

Helson, R., and R. S. Crutchfield. 1970. Mathematicians: The creative researcher and the average Ph.D. *J. Consulting and Clinical Psychology* 34:250-57.

Holland, J. L., and J. M. Richards, Jr. 1965. Academic and nonacademic accomplishment: Correlated or uncorrelated? *J. Educational Psychology* 56:165-74.

Hoyt, D. P. 1965. The relationship between college grades and adult achievement: A review of the literature. *ACT Research Report* No. 7. Iowa City: American College Testing Program.

Hoyt, D. P. 1966. College grades and adult accomplishment: A review of research. *Educational Record* (Winter) 70-75.

Hudson, L. 1960. Degree class and attainment in scientific research. *Brit. J. Psychology* 51:67-73.

Jencks, C., and D. Riesman. 1968. *The Academic Revolution.* Garden City: Doubleday.

Jordan, L. A. 1975. Use of canonical analysis in Cropley's "A five-year longitudinal study of the validity of creativity tests." *Developmental Psychology* 11:1-3.

Kogan, N., and E. Pankove. 1974. Long-term predictive validity of divergent-thinking tests: Some negative evidence. *J. Educational Psychology* 66:802-10.

MacKinnon, D. W. 1968. Selecting students with creative potential. In *The Creative College Student: An Unmet Challenge*, ed. P. Heist. San Francisco: Jossey-Bass. Pp. 101-16.

Marston, A. R. 1971. It is time to reconsider the Graduate Record Examination. *Am. Psychologist* 26:653-55.

Maxey, E. J., and V. J. Ormsby. 1971. The accuracy of self-report information collected on the ACT test battery: High school grades and items of nonacademic achievement. *ACT Research Report* No. 45. Iowa City: American College Testing Program.

Mednick, M. T. 1963. Research creativity in psychology graduate students. *J. Consulting Psychology* 27:265-66.

Mischel, W. 1972. Direct versus indirect personality assessment: Evidence and implications. *J. Consulting and Clinical Psychology* 38:319-24.

Morrison, J. 1963. The comparative effectiveness of intellective and nonintellective measures in the prediction of the completion of a major in theater arts. *Educational and Psychological Measurement* 23:827-30.

Munday, L. A., and J. C. Davis. 1974. Varieties of accomplishment after college: Perspectives on the meaning of academic talent. *ACT Research Report* No. 62. Iowa City: American College Testing Program.

Parloff, M. B., L. Datta, M. Kleman, and J. H. Handlon. 1968. Personality characteristics which differentiate creative male adolescents and adults. *J. Personality* 36:528-52.

Richards, J. M., Jr., J. L. Holland, and S. W. Lutz. 1967. Prediction of student accomplishment in college. *J. Educational Psychology* 58:343-55.

Singer, D. L., and M. B. Whiton. 1971. Ideational creativity and expressive aspects of human figure drawing in kindergarten-age children. *Developmental Psychology* 4:366-69.

Taylor, D. W. 1958. Variables related to creativity and productivity among men in two research laboratories. In *The Second (1957) University of Utah Research Conference on the Identification of Creative Scientific Talent*, ed. C. W. Taylor. Salt Lake City: University of Utah Press. Pp. 20-54.

Tyler, L. E. 1965. *The Psychology of Human Differences.* 3rd ed. NY: Appleton-Century-Crofts.

Wallach, M. A. 1967. Creativity and the expression of possibilities. In *Creativity and Learning*, ed. J. Kagan. Boston: Houghton Mifflin. Pp. 36-57.

Wallach, M. A. 1968. Review of the *Torrance Tests of Creative Thinking. Am. Educational Res. J.* 5:272-81.

Wallach, M. A. 1970. Creativity. In *Carmichael's Manual of Child Psychology*, 3rd ed., P. H. Mussen, ed. N.Y.: Wiley. Pp. 1211-72.

Wallach, M. A. 1971a. *The Intelligence/Creativity Distinction.* Morristown, N.J.: General Learning Press.

Wallach, M. A. 1971b. Intelligence tests, academic achievement and creativity. *Impact of Science on Society* 21:333-45.

Wallach, M. A. In press. The psychology of talent and graduate education. In *Cognitive Styles and Creativity in Higher Education*, ed. S. Messick. San Francisco: Jossey-Bass.

Wallach, M. A., and N. Kogan. 1965. *Modes of Thinking in Young Children: A Study of the Creativity-Intelligence Distinction.* N.Y.: Holt, Rinehart and Winston.

Wallach, M. A., and M. I. Leggett. 1972. Testing the hypothesis that a person will be consistent: Stylistic consistency versus situational specificity in size of children's drawings. *J. Personality* 40:309-30.

Wallach, M. A., and C. W. Wing, Jr. 1969. *The Talented Student: A Validation of the Creativity-Intelligence Distinction.* N.Y.: Holt, Rinehart and Winston.

Wallbrown, F. H., and C. R. Huelsman, Jr. 1975. The validity of the Wallach-Kogan creativity operations for inner-city children in two areas of visual art. *J. Personality* 43:109-26.

Weitzman, R. A. 1972. It is time to reconsider the GRE: A reply to Marston. *Am. Psychologist* 27:236-38.

Wing, C. W., Jr., and M. A. Wallach. 1971. *College Admissions and the Psychology of Talent.* N.Y.: Holt, Rinehart and Winston.

Chairman **Perkins**. Do you have something to put in the record?

Mr. **Drexeler**. Yes. I referred to an article in Newsday before and there is also one in the New York Times.

I thank you very much.

Chairman **Perkins**. I thank you very much.

[Whereupon, at 2 p.m. the subcommittee adjourned.]

APPENDIX

Prepared Statement of Gordon M. Ambach, President, the University of the State of New York and Commissioner of Education

I am Gordon M. Ambach, Commissioner of Education and President of the University of the State of New York. I appreciate this opportunity to submit a statement for the record on the subject of admissions testing legislation.

The New York State Education Department is often viewed as the nation's leading publisher of "open" tests, which is the label currently attached to tests that become available to the examinees and the general public after the tests have been administered. Open testing has characterized our oldest testing program, the Regents examination program, since its origin in 1865. During the current school year, the New York State Education Department will publish 36 Regents examinations, 16 other high school achievement tests, and 10 competency tests — a total of 62 examinations — that are completely open and available to anyone who wants them. In addition, 8 other tests will be administered, as a part of our pupil evaluation and competency programs, that are "semi-open" in the sense that they are available for inspection in thousands of school buildings throughout the state. All of these tests are published for use in New York State schools, as a means of maintaining academic standards and providing grist for the mill of accountability.

Nevertheless, we find it necessary to oppose the enactment of HR 4949, the Educational Testing Act of 1979, as drafted presently. The essence of the bill is disclosure of tests. While it may be desirable to have such disclosure for those who would want to see their papers and results, the cost to all of those tested because of the particular disclosure provision in HR 4949 seems to outweigh the benefits. The bill is concerned with all tests which are used for postsecondary or professional school admissions or tests used for preliminary preparation for such tests. Such tests will include the Scholastic Aptitude Test, the American College Testing Program assessment, the Graduate Record Examinations, the Miller Analogies Test, the Medical College Admission Test, and the Law School Admission Tests. As we are learning in New York, the law may include tests not designed for or, in some cases, inappropriately being used for admissions. An example is a portion of the medical licensing examination which is used for admission of many foreign trained medical students. It is conceivable that HR 4949 could be interpreted to apply to our Regents examinations because postsecondary institutions use them as part of the admissions process. Although these tests are already open and offered free of charge to students, other provisions of the law would create a severe financial burden for the state and no benefit to the students.

As I indicated, the New York State Education Department publishes many open tests each year. They are educational achievement tests, and they are open because they serve as a means of communication between the Education Department and the classroom. They tell teachers and students exactly what outcomes of instruction are important, and they accomplish this task much more effectively than any of the other forms of communication that we use, including the publication of syllabuses, course outlines, study guides, bulletins, and newsletters. The educational benefits of having open Regents examinations and competency tests clearly outweigh the drawbacks of producing dozens of new examinations each year.

We have also had experience with secure tests. For many years we administered a secure examination for the purpose of selecting winners of the more than 18,000 Regents college scholarships that are awarded annually to high school seniors. Currently, the Department's College Proficiency Examination Program uses secure examinations as a service to individuals who wish to earn college credit for knowledge and skills acquired outside of ivy-covered walls. Many of our professional licensing examinations are also secure. Because of our need to use secure tests for some purposes, I understand many of the problems facing the test publishers and sponsors affected by the bill.

Looking only at the cost implications of H.R. 4949, there can be no question that the publishers of admissions tests would be required to increase their expenditures for test development in order to maintain their current testing schedules. Assuming that about 20% of the items used on each secure tests are new items, the cost of developing open tests would be five times greater than the cost of developing secure tests. Assuming again that the actual total direct and indirect cost of secure test development represents 25% of a publisher's total expenditures, the total expenditures would double with open testing. Using a different set of assumptions, one may forecast a smaller or larger impact on the publishers of admissions tests, but it is difficult to avoid the conclusion that the impact will be significant. The greatest effect will be on publishers of tests taken by relatively small groups of applicants. It is a realistic possibility that some of these tests may be forced out of existence because the applicants will be unable to absorb the cost. In particular, I am thinking of tests which are used for admission to highly competitive graduate level professional programs. Which applicants would be hurt the most if one of these tests was withdrawn? I doubt that it would hurt the elite students from the most prestigious colleges. The lack of a supporting test score will have the greatest impact on the adult learner with a "stale" undergraduate background, the economically disadvantaged student who was forced to attend a less prestigious institution, and the student who obtained lower grades in the first year or two of college because of a disadvantaged background. In other words, the law may hurt some of the students it was designed to help.

Unfortunately, there is another step that publishers could take to minimize the impact of the proposed legislation. They could simply reduce the number of tests administered. For example, instead of offering the SAT on seven Saturdays and seven Sundays, its publisher could administer the test only twice a year, thereby reducing the optional testing sites and dates for the student. The greatest impact will be on applicants who are given the opportunit to take special administrations of these major tests because they cannot attend the regular administrations for various reasons including illness, physical handicaps or religious beliefs. We estimate that special administrations of some major tests are actually given on nearly one hundred different dates during the year. No publisher can be expected to continue this practice under the proposed law.

Our general concern for New York State residents is heightened by the fact that six of the admissions tests affected by the proposed legislation are use as the basis for awarding scholarships and fellowships in eight different scholarship and fellowship programs administered by the State Education Department. The major argument for using these tests is to avoid the needless duplication of testing which would result from the use of Department-prepared examinations, to the extent that our own tests would parallel the various admissions tests. Thus, the negative impact of the proposed legislation would be most acute for scholarship and fellowship candidates, many of whom have the least ability to pay the increased fees and travel to other locations to take the examinations.

The use of admissions tests for scholarship and fellowship purposes introduces an additional concern and potential source of increased fees. As a convenience to candidates, they may submit scores obtained on any testing date prior to the filing deadline for the scholarship or fellowship application. This avoids many administrative problems associated with a single testing date. At the same time, it creates a need for test scores that have identical scale values regardless of the form that is taken.

A major criterion for admissions tests is that the scores be comparable between different forms of the test. This requires a benchmark or standard for equating. There are two basic strategies used for this process. The first involves simultaneous administration of a new and old form. This commonly used alternative will be eliminated by the proposed HR.4949 bill. The second strategy involves the use of a few equating questions on each new form. Although these equating questions may be kept secure, this is not a simple solution to a complicated problem. I urge you to reconsider the equating problems which have been cited in other presentations.

Many of the problems of equating are made acute by the short time period imposed before implementation. Test development is a long and complicated process. Many of the tests to be administered several months from now have already been developed. Many of these have been developed using the first strategy mentioned above. Can they be changed within a few weeks to provide accurate and consistent scores using a different strategy? I doubt it. If a new and unproven equating procedure is used, only the test candidates will suffer if it fails.

There is also a potential threat to the inclusion of non-operational items on a test form solely for use in equating and pretesting. In the Scholastic Aptitude Test, examinees spend one-sixth of their testing time answering questions that have no bearing on their scores, and they have no idea which items these are. If this test becomes open, one can anticipate considerable pressure on its publisher to eliminate the non-operational items. This will force the publisher to devise new, more costly, and less accurate methods for equating test forms and for pretesting questions.

A test, like any other product, may contain faults. No reputable test publisher would claim that its test is perfect. The dependability of a complex industrial product is based in part on the use of tried and proven components. An examination composed of previously used and proven questions is no different. If all of these questions must be replaced by new and untried questions for each administration, the quality of the product will decrease.

With reference to the benefits that are claimed for HR 4949, it may be worthwhile to examine them more closely. One benefit is that it would make tests available for all to inspect, evaluate and administer in independent research. As it now stands, however, some publishers have released forms of their tests for these very purposes. Further, this benefit could be realized by simply requiring all publishers to release one or two forms of their secure tests. It is unlikely that this would force publishers to redesign their entire systems of test development and increase fees or decrease services.

A second benefit claimed for H.R. 4949 is that it would allow examinees to review their answer papers in order to verify the accuracy of the scoring and challenge any questions that are poorly written or incorrectly keyed. Assuming that this is necessary or desirable, it, too, could be accomplished by less radical surgery. Copies of answer papers and scoring keys could be furnished and the publishers could be required to designate certain locations in which examinees could review question keys under secure conditions. This is the procedure followed with New York State Civil Service Examinations, and it works well without any adverse effects upon the cost and quality of the examinations.

In sum, while the State Education Department agrees with the intention to assure that the purposes, procedures, and use of admission tests are made available to the public, H.R. 4949, as drafted, is not an effective means toward that end. Since H.R. 4949 is patterned after the New York State statute, we urge a delay in the enactment of any Federal testing legislation until the full impact and ramifications of the New York State statute are realized. In this way, any adverse effects from the New York State experience can thereby be prevented in future Federal legislation. If a delay is not possible, we urge that alternatives be developed which will not have adverse effects upon the cost, quality and accessibility of the examinations.

THE UNIVERSITY OF THE STATE OF NEW YORK
THE STATE EDUCATION DEPARTMENT
OFFICE OF THE PRESIDENT OF THE UNIVERSITY
AND COMMISSIONER OF EDUCATION
ALBANY, NEW YORK 12234

October 17, 1979

Dear Mr. Chairman:

On August 13, 1979, I submitted a statement for the record opposing enactment of HR 4949 the "Educational Testing Act of 1979."

Governor Carey on enacting the New York State Standardized Testing Law (Education Law Article 7-A as enacted by Chapter 672 of the Laws of 1979) asked the Department to prepare a report on the impact of this law. I have recently submitted to Governor Carey our initial report and wish to include it as a part of my earlier statement submitted to you for the record. I hope this will be helpful to you as you further deliberate HR 4949.

If I may be of any further assistance, please do not hesitate to call upon me.

Sincerely,

Gordon M. Ambach

Honorable Carl D. Perkins
Chairman
Subcommittee on Elementary, Secondary
and Vocational Education
Committee on Education and Labor
Washington, DC 20515

cc Senator Javits
Representatives Biaggi, Weiss, Peyser, Chisholm and Goodling

THE UNIVERSITY OF THE STATE OF NEW YORK
THE STATE EDUCATION DEPARTMENT
OFFICE OF THE PRESIDENT OF THE UNIVERSITY
AND COMMISSIONER OF EDUCATION
ALBANY, NEW YORK 12234

October 16, 1979

Dear Governor Carey:

I am pleased to transmit an initial statement concerning the anticipated impact of Chapter 7A of the Education Law. This report is submitted in accordance with the request in your approval message of July 13, 1979.

I will shortly send this report to New York colleges and universities asking how they plan to deal with the anticipated changes occasioned by the legislation. An analysis of the responses from the colleges will furnish information for a subsequent report.

In November, I will bring to the Regents proposed amendments to the Regulations of the Commissioner of Education to implement the Law. Both the proposed Regulations and commonly asked questions and answers about them are included as part of this impact report.

Sincerely,

Gordon M. Ambach

The Honorable Hugh L. Carey
The Governor of the State of New York
State Capitol
Albany, New York

THE UNIVERSITY OF THE STATE OF NEW YORK
THE STATE EDUCATION DEPARTMENT
ALBANY, NEW YORK 12234

Initial Impact Report
# of the
Admission Testing Legislation

Introduction

Debate over the proposed admission testing legislation generated numerous statements of support and opposition. In his memorandum of approval of the bill, the Governor acknowledged these conflicting statements and asked the Commissioner of Education to report on the impact of the legislation and on any administrative problems caused by its implementation. This report outlines the potential impact of actions being considered by the test agencies. The full impact of the legislation will not be known for many months and a more complete report will be generated after the major test agencies have completed their plans.

The impact of the admission testing legislation is influenced by the decisions of individual test sponsors. This report is divided into several sections. A brief background discussion is followed by a discussion of the tests being withdrawn from use in the State. The probable reasons for withdrawal of these tests as well as the impact on students are included. Next, the reduction in services for other test programs is described in terms of its impact on candidates. The last sections of this report discuss technical and administrative problems, one of which may have severe impact on testing during the second half of the 1979-80 academic year.

Background

On August 1, a few days after the admission testing bill was signed, the Education Department mailed a copy of the Law to over sixty test sponsors and publishers. A covering letter (Attachment 1) asked these agencies to identify the tests which come under provisions of the Law. The test agencies were also asked to provide their plans for testing after the effective date. As a result of this letter, the names of twenty-six admissions testing programs have been identified. In most cases a program consists of only a single test, however, in some cases a program contains several individual tests.

Although the names of admission testing programs were provided, the sponsors almost uniformly stated that it was premature for them to make plans for tests to be given after the effective date of the legislation. They expressed concern about the ambiguity of some sections of the Law and the possible consequences of various interpretations of the Law. Meetings and telephone conversations were held with individual sponsors to discuss the interpretation of the Law. The sponsors wanted all aspects of the Law to be defined by Commissioner's Regulations. The Department takes the position that interpretation of the law can only be made by the courts. The Department has expressed its opinion about various sections of the Law in a series of questions and answers. These opinions as well as proposed Commissioner's Regulations will be circulated within a few days. Copies of these two documents are included as Attachments 2 and 3.

Most sponsors now have provided the Department with an idea of the testing plans they are considering. Although the plans are tentative, they provide a basis for reporting on the probable impact of the law. Attachment 4 lists the various admission testing programs identified by the sponsors. These programs are divided between those which are expected to continue providing some services in New York State and those which are expected to withdraw services. In reviewing the list, it is critical that two important considerations be kept in mind. First, the list includes only tests identified by the sponsors. Other testing programs probably will be added to the list during the next few months. Second, the classification of a testing program into one of the two basic groups must be considered tentative. The sponsors of all of these tests appear to be working to overcome obstacles which are forcing curtailment or elimination of services in New York State.

TESTS TO BE WITHDRAWN FROM USE IN NEW YORK STATE

A review of Attachment 4 reveals that twenty of the twenty-six test programs may be withdrawn from use in New York State. Most of these tests are given to small numbers of persons compared with the large number of candidates taking the Scholastic Aptitude Test and the ACT Assessment Program. Small numbers of candidates is the reason cited by the sponsors for withdrawal of most of these tests. In a few other cases the flexibility of test administration procedures was cited as the major obstacle facing the sponsors.

The actual development cost of a new test form is relatively independent of the number of candidates to be tested. Although this cost may vary considerably, $25,000 is a conservative estimate of development cost for many half-day tests. This figure includes the cost of question writing, review, and validation; it does not include other costs such as test administration and overhead. These development and administrative costs are borne by the candidates taking the test. When the number of candidates is small, the cost per candidate is high unless it is possible to reduce the total cost by the reuse of individual questions. Under the new legislation, the release of questions after use will require development of totally new forms for each administration. Such development will be realistically impossible for the sponsors of many of these tests. At least six of the tests are administered to less than 500 New York State candidates each year. One sponsor provided the example of its graduate school admissions test which is administered three times each year to a total of 300 candidates in New York State. Using the estimate of $25,000 to simply develop each form, the sponsor stated that the total development cost would be $75,000 each year or $250 per candidate. Reducing the number of administrations to once each year would still result in a development cost in excess of $80 per candidate. On this financial basis alone the sponsor sees no alternative other than to withdraw the test from use in New York.

Some of the other tests in this group are given on locally scheduled dates set by independent administrators or individual institutions. Although a few of these tests are given to large numbers of candidates, the problems facing sponsors are illustrated in the following statement submitted by one sponsor:

> ..."The problems of using this flexible administration procedure under the LaValle Act relate to its requirement that after a single person is tested in any New York State Center in 1980 (or in an out-of-state Center but with results reported to a New York institution) we would be obliged to make the test content public after thirty days, with the resulting loss of test security, and need to create new secure forms.
>
> For us to change from this flexible arrangement to a fixed-testing-date plan in New York State only, would require the determination of two to four dates and negotiations to set up examining sites, since some existing Centers may not be interested in, or capable of, changing their style of operation. We would have to alter our methods of scoring and reporting of results. Whereas at present the tests are scored and initial reports are sent out by the local examining Center, the change for New York would require a scoring and reporting system operating out of our offices."

Many of the administrative problems alluded to in this statement could be overcome. To do so would require time, however, and changes would be made at the expense of the very flexibility which has made the tests to useful to the colleges.

Impact of Tests to be Withdrawn

If only a majority of the tests in this group are actually withdrawn, there will still be a clear impact on the State.

Sixteen of the twenty tests are designed for use by health related programs. This fact, by itself, is of major concern. The greatest negative impact will undoubtedly result from withdrawal of tests used by programs with highly competitive admissions procedures. Loss of the tests will not hurt the elite students from the most prestigious colleges. As the dean of one medical school reported, it is easy to pick only the top students from the top undergraduate programs. His school, and other, also review the applications of many other individuals who would not have been considered except for their good performance on the admissions test. This group often includes the economically disadvantaged student who was forced to attend a less prestigious college or the student who obtained lower grades in the first year or two of college because of a disadvantaged background. In some programs this group includes the adult learner with a "dated" undergraduate background. The percentage of students admitted from this group may be no more than ten to twenty percent, however, the numbers are still significant. Although specific data is unavailable, this group appears to contain a high proportion of minorities and economically disadvantaged students. During the last few years considerable effort has been made to increase the representation of minorities within the professions. The loss of the admission test as an independent measure of ability presents a partial roadblock to correction of this inequity. The disadvantaged student with superior undergraduate grades is likely to be admitted even with lower admission test scores. On the other hand, the student with equal ability, but less impressive credentials may never be considered without the benefit of a test score.

Less significance will result from withdrawal of the tests designed for use by practical nursing programs. These programs are able to accommodate most applicants who have achieved satisfactory high school grades. In most other health related programs, the competition for admission is keen.

The Regents Scholarship Program for Medicine, Dentistry and Optometry

Each year the Regents award a total of 135 scholarships in medicine, dentistry and optometry on a competitive basis as required by Education Law. Medical and dental scholarships were originally based on the results of a Department developed test designed specifically for this purpose. However, this test was eliminated by the 1977-78 Budget as was the test used for undergraduate scholarship awards. The admission tests used by medical and dental schools were then accepted as the alternative for determining scholarships. In 1977, optometry scholarships were created by the legislature and the Optometry College Admission Test was selected as the only appropriate test to use in making awards.

The sponsors of these three tests have informed the Department that the tests may be withdrawn from use in the State. If this occurs, the Regents will have no appropriate method of making the awards.

TEST PROGRAMS TO BE CONTINUED IN NEW YORK STATE

At least six admission testing programs are expected to continue serving New York State students and institutions. In all cases, however, services are expected to be reduced. The following is a brief description of the tentative plans which have been provided by the sponsors.

ACT Assessment Program

The sponsor has provided no specific plan for reductions in the number of regular administration dates, however, continued availability with some decrease in services is anticipated. Telephone conversations indicate that special administrations probably will be sharply reduced or eliminated. Most significant in terms of numbers of candidates is the probable elimination of residual (on-campus) testing. Secure forms of the test are currently available to colleges for use in evaluating applicants who did not take a regular form of the test. Approximately 2,200 students were tested in this way during the last testing year; many, if not the majority, were adults who needed the scores for placement rather than admission. Over 300 other students were tested on non-Saturday test dates (usually the following Sunday) due to religious reasons. To avoid any questions over the validity of scores caused by a compromise in security, these students are given a different form of the test. In the future, the number of special non-Saturday test dates may be drastically reduced. Handicapped students are now permitted to take a form of the test on virtually any day in the year. This practice probably will be eliminated and these candidates will be expected to take the test on the regularly scheduled administration dates.

Graduate Management Admissions Test

This test is normally administered on a Saturday and on the following Monday in the months of October, January, March and July. The January and March 1980 dates will be continued but the Monday administration in July will probably be cancelled. This will affect approximately 140 sabbath observers. While the sponsor has not determined its New York testing schedule for the following year, it states "...While it may be possible to maintain the same number of Saturday test dates, it may be necessary to reduce the number of times the test is offered on Monday." The sponsor had planned to introduce special forms (large print, Braille and cassette) of the test in July 1980. Because of the time and money required to produce these forms, their introduction will be postponed. Thereafter, testing using these formats may be limited to one or two dates a year. Make-up testing for administrations cancelled due to unforeseen emergencies will be available only when the same form of the test given on the regular date at other locations can be reused before public disclosure.

Graduate Record Examination Program

The Graduate Record Examination Program consists of one general Aptitude Test and numerous Advanced Tests in different subject areas. The Advanced Tests were specifically excluded from the disclosure provisions of the Law and it is unlikely that their availability will be reduced. The availability of the Aptitude Test probably will be reduced from the current six times per year to four or possibly fewer times per year. Also, Monday administrations for sabbath observers which are now held after five of the six Saturday dates will be reduced. It is likely that the sponsor may cancel the administrations already scheduled for Monday in

January, Saturday in February (no Monday date was scheduled) and both the Saturday and Monday dates in June. The only remaining test dates will be the Saturday administration in January and the Saturday and Monday administrations in April. Approximately 5,600 potential graduate students will be affected by these probable cancellations. Further, the sponsor states that:

> "Until the effective date of the legislation, it will be possible to continue past GRE practice of reusing secure forms (or editions) of the Aptitude Test. This flexibility has permitted the development of a variety of special services to meet the needs of graduate institutions and their applicants who are required to submit GRE scores. The requirements of the La Valle law make disclosure of each test after small administrations prohibitively expensive because of the increase in the number of new test forms that would be required to ensure secure testing conditions..... Currently, handicapped students may take the Aptitude Test on any scheduled administration date (national or "special") and may use, as appropriate, tests presented in large print, braille, or on cassette tapes. Because of the high costs of producing test versions in braille and on cassettes, these two options will not be provided in New York in April 1980. Thereafter, testing using either of these formats, and possibly large print as well, will be restricted, if available at all, to one or two dates a year."

In addition, the Aptitude Test is available to military personnel through the Defense Activity for Non-traditional Education Supports. Because many of the military personnel might want results sent into New York, the DANTES testing will be temporarily discontinued on January 1, 1980, except for testing during the week following the April national test date.

Law School Admissions Test

The sponsor has conveyed its clear desire to continue providing services within the State. It has not, however, determined the specific actions it will take for test dates already scheduled and for dates to be scheduled in subsequent test years. In developing its schedule, the sponsor has particular concern over how to accommodate the relatively large number of sabbath observers who require special testing dates.

Preliminary Scholastic Aptitude Test/National Merit Scholarship Qualifying Test

The first administration of this test subject to provisions of the Law would be October 1980. Although regular administrations on a Tuesday and Saturday in October are still envisioned, it is likely that arrangements for alternate testing will have to be discontinued.

Scholastic Aptitude Test and Achievement Tests

The College Board Achievement Tests as well as the GRE Advanced Tests mentioned earlier are excluded from disclosure provisions of the law. For this reason, no major long term test administration changes are anticipated. The Scholastic Aptitude Test (SAT) on the other hand, is subject to all provisions of the law. On October 10, the College Board, sponsor of the test, announced cancellation of the January administration and two other scheduled Sunday administrations. The January Achievement Test administration was also cancelled because of the high cost of opening test centers for a comparatively small volume of candidates. As many as five of the Sunday administrations, affecting 1,800 sabbath observers, may be cancelled in 1980. In addition, the sponsor is considering the cancellation of all January and November dates which would adversely affect an additional 28,000 students.

A very limited number of forms of a Spanish version of the SAT (Prueba de Aptitud Academica) are developed in Puerto Rico. The sponsor does not feel it is possible to identify even one form of the test for disclosure in New York. This decision will affect approximately 400 New York students.

Alternate and make-up testing as well as institutional administrations probably will be eliminated. The number of students affected may vary from about 350 to 750 each year.

"Current practice permits testing of handicapped students on virtually any date and at any place mutually convenient to the student and test administrator. Because of the high cost of converting tests into Braille, large type and on audio-cassettes, and the subsequent disclosure of each edition of the test, handicapped testing in New York State will be restricted to a very limited number of scheduled test dates." Approximately 400 New York students will be affected by this action.

"The tentative plans described above do not include reference to two other important factors on which the sponsors have not reached final decisions. The first is fee increases to cover the costs incurred in meeting the provisions of the law. All sponsors anticipate an increase in the test fees, however, no indication of the amount of change has been given. Thus it is impossible to predict the cost impact. The second concerns the current practice of providing statistical reports to institutions. This factor will be discussed in the following section on impact.

Impact of Tests Offered on a Reduced Schedule

The six testing programs mentioned above are taken by thousands of students each year. Most postsecondary institutions use the results as part of the admission selection process; the results are also used for several scholarship programs. For these reasons changes in availability which appear relatively minor still can have a serious impact on the large population of potential candidates.

Sabbath observers may be the specific group most adversely affected by the law. The regular administrations of these tests are given on Saturdays to avoid school, employment and religious conflicts for the vast majority of test candidates. Alternate test dates, usually on Sunday or Monday, have been provided after virtually every administration for those candidates whose religious beliefs preclude Saturday testing. The exact number of sabbath observers tested during the last year is believed to be between 4,000 and 5,000. Unfortunately, this very significant total appears small when distributed among the various administrations of the different tests. The reduction in the number of Sunday and Monday administrations is expected to be drastic. Some tests may be made available to sabbath observers on only one or two dates in the year. Candidates who miss this date or dates for legitimate reasons will have no opportunity to be tested in the reasonable future. These candidates will be at a disadvantage when applying to postsecondary institutions and they may be ineligible for scholarship competition.

Physically handicapped students are provided with various extra services including administrations held in hospitals and the candidates' homes on virtually any date. The sponsors which provide testing at these special locations have indicated that they will continue this practice but only on the regularly scheduled dates. The majority of physically handicapped candidates have visual handicaps which require special forms of the tests. Based on the tentative plans of the sponsors it appears that "large-type" editions of most tests will be generally available, but Braille and cassette formats will be available for only a limited number of test dates. The high cost and limited usage of these special forms is the reason given by the sponsors. This decision will affect over 500 handicapped candidates annually.

Other special test administrations are held annually for

3,000 to 4,000 candidates who could not be tested on the regular dates. Many of these candidates take the test through the post-secondary institution to which they have applied. This special service is expected to be eliminated by the sponsors who now make it available. The result is that some candidates may have their enrollment delayed for a semester or a year and other candidates may be denied the credit or advanced standing which the test would validate. This latter factor particularly affects candidates seeking undergraduate admission and placement.

All candidates are likely to feel some effects of the anticipated reduction in test administration dates. A reduction in the number of administrations will increase the number of candidates tested on the remaining dates. This will burden some administration centers and necessitate the need for new centers run by inexperienced personnel. Such necessity may in turn adversely affect the quality of the administrations until adequate test sites are located and personnel are trained. To help compensate for this potential problem, a few sponsors have indicated that it may be necessary to reduce special services. These services are likely to include the provisions for late registration and "walk-in" candidates.

All candidates may be indirectly affected by the sponsors action relative to the provision of the Law regarding disclosure of statistical reports. Institutions generally receive reports from the test sponsors summarizing test information about candidates applying to the institution and, in some cases, comparing these applicants to all applicants for other institutions combined. This type of information gives the institution an estimate of the quality of its applicants compared with the quality of applicants in general.

Test sponsors also provide validity studies used by the institutions to establish and refine their selection procedures. These various reports and studies are considered confidential documents for the individual institutions. The Law, however, requires that all statistical reports become public. To avoid being placed in a potential legal conflict between the Law and the institutions' rights, the test sponsors have indicated that they probably will stop routinely providing this service to the institutions. Although these data still could be generated upon request of the institution, very few institutions are likely to request this service since they will not want this information made public. In the long run, the loss of validity data will adversely affect the selection process and, thus, the applicants.

ADMINISTRATIVE PROBLEMS RESULTING FROM THE LAW

The Effective Date

Most admission tests are administered on dates during the school year. For this reason, in part, most test sponsors use an academic year rather than a calendar year. Registration materials are prepared in the spring for use by candidates planning to be tested on any of the dates scheduled during the following school year. These materials for the 1979-80 year have already been distributed to many prospective candidates through local high schools and colleges. The sponsors of several admission tests including the Scholastic Aptitude Test and the ACT Assessment Program report that several thousand candidates have already filed applications for testing dates after January 1, 1980. This places the sponsors in a serious legal dilemma.

The Law requires test agencies to provide specified information "... along with the registration form for the test..." Although the majority of this information is already provided, no registration booklet contains all the necessary information. Further, there is no guarantee that distribution of supplemental material would reach every prospective candidate who has already obtained the registration material. For those candidates who have already filed applications, it is clearly too late for test sponsors to comply with the provisions of the Law. The sponsors must make a difficult choice even if their long term intention is to comply fully with provisions of the Law. First, they can continue with administrations scheduled between January and June 1980, even though they are technically in violation of the law and subject to a $500 penalty for each offense. Second, they can cancel these administrations and schedule new dates after the registration materials have been revised. Last year approximately 75,000 candidates in New York State completed the January 1979 administration of the SAT. If the January 1980 administration had been held, the College Board would have been subject to a penalty in excess of seven million dollars. Although this is the most dramatic example, other sponsors face comparable problems.

Interpretation of the Law

Several different provisions of the Law have evoked major questions of interpretation. Test agencies and other interested parties have asked the Department to clarify and interpret the Law through Commissioner's Regulations. This use of Regulations is sometimes appropriate when the Commissioner is empowered by a law to determine violations. In this case, however, the Commissioner is empowered only to collect information which must be furnished by the test agencies. As a result, interpretation as well as the determination of violations must rest with the courts. Numerous lengthy court cases could evolve which would be detrimental to all parties. In all likelihood the test candidates would be the major victims of these delays.

Impact on the Department

The probable loss of the tests used to award scholarships in medicine, dentistry and optometry has already been cited as a problem.

The Law gives the Commissioner and the Education Department certain administrative responsibilities but no appropriation for this specific purpose. The Department believes that significant resources may be needed to meet its responsibilities under the new Law. Since the bill was signed, it has been necessary to have one management level employee devote virtually full time to the topic. In addition, numerous other devote virtually full time to the topic. In addition, numerous other management, professional and support staff have been involved on a part-time basis. This involvement is expected to increase until well after the effective date of the Law. Thereafter, the involvement of management level staff may decrease but it will be offset by the involvement of other staff. The cataloging of thousands of pages of statistical data will require the use of trained professional staff. This and other required related functions will require many hours of work. A specific example is the need to communicate with test candidates who disagree with the test agencies over the interpretation of test questions. Thousands of candidates are expected to file complaints with the test agencies over the questions and answers for each test. Most of these candidates will be informed by the test agencies that their complaints are without merit, and then many of the candidates will turn to the Department for assistance. Although not mandated by the Law, the Department believes it is obligated to communicate with the candidates and serve as an ombudsman.

The Department has an obligation to monitor compliance with the Law. Apparent violations will be brought to the attention of the test sponsor and, if necessary, the Attorney General. As a first step, all postsecondary and professional schools in the State will be asked to identify all tests used for admissions purposes. This is likely to result in the identification of other tests which are actually being used for admission selection purposes. It then will be possible to alert students, institutions and test agencies that the Law applies to these tests.

Other types of involvement and intervention may also be needed.

SUMMARY

Twenty-six postsecondary school admission testing programs used in New York State have been identified by the test agencies. Based on very tentative information available by October 15, about twenty of these programs may be withdrawn from use in the State. Postsecondary institutions will be surveyed within the next few weeks to help determine the current use of these test programs and to identify other testing programs being used by the institutions for admission selection purposes. It is likely that other tests, not designed for admissions purposes, are actually being used in this way by the institutions.

Five of the six programs expected to continue in the State are published by the Educational Testing Service for various sponsors including the College Board. The sixth is the ACT Assessment Program sponsored by The American College Testing Program. The sponsors of these six test programs have all indicated their desire to continue testing, however, they all expect to reduce services. The number of regular administrations is expected to be reduced somewhat and special testing will be curtailed sharply. This step is likely to have the greatest effect on the physically handicapped and sabbath observers.

About half of the twenty test programs which may be withdrawn have very limited usage in the State and their elimination may not produce a severe impact. The other half, however, include several tests used for admission to highly competitive health related programs such as medicine and dentistry. Also to be withdrawn are tests such as the Miller Analogies Test which is used by many graduate level programs in the State. Withdrawal of these tests is expected to have the greatest impact on applicants with prior academic backgrounds which are questionable or unfamiliar to the institution.

THE UNIVERSITY OF THE STATE OF NEW YORK
THE STATE EDUCATION DEPARTMENT
CULTURAL EDUCATION CENTER
ALBANY, NEW YORK 12230

ASSISTANT COMMISSIONER
FOR HIGHER EDUCATION SERVICES

August 1, 1979

Dear Test Publisher/Sponsor:

As you may know, legislation has recently been enacted in New York which requires filing with the Commissioner of Education and disclosure to the public of certain information concerning tests used in the process of the admission of students to post-secondary and professional schools in New York. A copy of the law, which was enacted on July 13, 1979 and becomes effective January 1, 1980, is attached for your information.

The provisions of the new law apply to any test which is given at the expense of the test subject and is designed for use or in fact used in the process of selection for post-secondary or professional school admissions, where the test is administered in New York State or the results of the test are provided to institutions located within the State.

The New York State Education Department is compiling a list of the tests which are subject to the law indicating the planned administration dates for each test during the period January 1, 1980 through December 31, 1980. This information will be used to plan our administration of the law, and to prepare a report to the Governor and the Legislature on the anticipated impact of the law.

May I ask that you provide us with a list of all tests developed, sponsored or administered by your agency for use by post-secondary or professional schools in the process of selection of students for admissions. Please indicate which of those tests will be administered in New York State in 1980 or, without regard to the place of administration, those tests the results of which will be provided to institutions located in New York State. Will you also indicate your plans and schedule for the administration of such tests in 1980. We will greatly appreciate your transmitting this information by August 15, 1979.

The information requested above, and all information and materials which must be filed with the State Education Department in accordance with the new law should be addressed as follows:

Mr. David R. Bower
Administrator of Postsecondary
Test Programs
New York State Education Department
Room 5D28
Cultural Education Center
Albany, New York 12230

Thank you for your assistance and cooperation.

Sincerely,

Alvin P. Lierheimer

Attachment

STATE OF NEW YORK

S. 5200—A A. 7668—A
Cal. No. 1215

1979-1980 Regular Sessions

SENATE–ASSEMBLY

April 26, 1979

IN SENATE—Introduced by Sens. LaVALLE, ACKERMAN, BABBUSH, BARTOSIEWICZ, BEATTY, BERMAN, BERNSTEIN, BRUNO, CAEMMERER, CONNOR, FLYNN, GALIBER, GOODHUE, HALPERIN, LACK, LEICHTER, MARKOWITZ, McCALL, MEGA, MENDEZ, OWENS, PADAVAN, PISANI, PRESENT, RUIZ, SOLOMON, TAURIELLO, TRUNZO, VOLKER, WINIKOW—read twice and ordered printed, and when printed to be committed to the Committee on Education—reported favorably from said committee and committed to the Committee on Higher Education—reported favorably from said committee, ordered to first and second report, ordered to a third reading, amended and ordered reprinted, retaining its place in the order of third reading

IN ASSEMBLY—Introduced by COMMITTEE ON RULES—(at request of M. of A. Vann, LaFayette, Bianchi, Boyland, Cochrane, Cooperman, Eve, Farrell, Flanagan, Fortune, Gottfried, Grannis, Griffith, Harenberg, Hinchey, Hirsch, Hochbrueckner, Howard, Jacobs, Jenkins, Koppell, Lentol, Lewis, Lipschutz, McCabe, G. W. Miller, Nine, Passannante, Pesce, Proud, Serrano, Siegel, E. C. Sullivan, Virgilio, D. B. Walsh, Yevoli, Zagame)—read once and referred to the Committee on Higher Education—reported and referred to the Committee on Ways and Means—committee discharged, bill amended, ordered reprinted as amended and recommitted to said committee

AN ACT to amend the education law, in relation to standardized testing

The People of the State of New York, represented in Senate and Assembly, do enact as follows:

Section 1. The education law is amended by adding a new article seven-A to read as follows:

ARTICLE 7-A
STANDARDIZED TESTING

Section 340. Definitions.
341. Background reports and statistical data.
342. Disclosure of test contents.
343. Notice.
344. Disclosure of test scores.
345. Regulations.

EXPLANATION — Matter in *italics* is new; matter in brackets [ ] is old law to be omitted.

[illegible]

346. Violations.
347. Severability.

§ 340. Definitions. As used in this article: 1. "Standardized test" or "test" means any test that is given at the expense of the test subject and designed for use or used in the process of selection for post-secondary or professional school admissions. Such tests shall include, but are not limited to, the Preliminary Scholastic Aptitude Test, Scholastic Aptitude Test, ACT Assessment, Graduate Record Examination, Medical College Admission Test, Law School Admission Test, Dental Admission Testing Program, Graduate Management Admission Test, and Miller Analogies Test. This article shall not apply to any state, federal, or local civil service test, any test designed and used solely for non-admission placement or credit-by-examination, or any test developed and administered by an individual school or institution for its own purposes only.

2. "Commissioner" means the commissioner of education of the state of New York.

3. "Test subject" shall mean an individual to whom a test is administered.

4. "Test agency" shall mean any organization, association, corporation, partnership, or individual or person that develops, sponsors or administers a test.

§ 341. Background reports and statistical data. 1. Whenever any test agency prepares, causes to have prepared or provides the data which are used in any study, evaluation or statistical report pertaining to a test, such study, evaluation or report shall be filed with the commissioner.

2. If any reports or other documents submitted pursuant to this section contains information identifiable with any individual test subject, such information shall be deleted or obliterated prior to submission.

3. All data, reports or other documents submitted pursuant to this section shall be public records.

§ 342. Disclosure of test contents. 1. Within thirty days after the results of any standardized test are released, the test agency shall file or cause to be filed in the office of the commissioner:

a. a copy of all test questions used in calculating the test subject's raw score;

b. the corresponding acceptable answers to those questions; and

c. all rules for transferring raw scores into those scores reported to the test subject together with an explanation of such rules.

2. After the test has been filed with the commissioner, and upon the request of the test subject, the test agency shall send to the test subject:

a. a copy of the test questions used in determining the subject's raw score;

b. the subject's individual answer sheet together with a copy of the correct answer sheet to the same test with questions counting toward the subject's raw score so marked; and

c. a statement of the raw score used to calculate the scores already sent to the subject, provided that such request has been made within ninety days of the release of the test score to the test subject.

The agency may charge a nominal fee for sending out such information, not to exceed the direct cost of providing the information.

3. This section shall not apply to College Board Achievement Tests or GRE Advanced Tests.

4. Documents submitted to the commissioner pursuant to this section shall be public records.

§ 343. Notice. 1. Each test agency shall provide, along with the registration form for a test, the following information:

a. The purposes for which the test is constructed and is intended to be used.

b. The subject matters included on such test and the knowledge and skills which the test purports to measure.

c. Statements designed to provide information for interpreting test results, including but not limited to, explanations of the test score scale, the standard error of measurement of the test, and a list of available correlations between test scores and grades, successful completion of a course of study and parental income.

d. How the test scores will be reported, whether the raw test scores will be altered in any way before being reported to the test subject and whether and how the test agency will use the test score in raw or transformed form by itself or together with any other information about the test subject to predict in any way the subject's future academic performance for any post secondary educational institution.

e. A complete description of any promises or covenants that the test agency makes to the test subject with regard to accuracy of scoring, timely forwarding of information, policies for notifying test subjects regarding inaccuracies in scoring or score reporting and privacy of information relating to the test subject.

f. Whether or not the test scores are the property of the test subject, how long they will be retained by the test agency, and policies regarding storage, disposal and future use of test score data.

2. Any institution which is a test score recipient shall be provided with the information specified in this section. The test agency shall provide such information prior to or coincident with the first reporting of a test score or scores to a recipient institution. Such institution shall be encouraged to provide interpretive processing by qualified personnel where such personnel are available.

§ 344. Disclosure of test scores. The score of any test subject shall not be released or disclosed by the test agency to any person, organization, corporation, association, college, university, or governmental agency or subdivision unless specifically authorized by the test subject. A test agency may, however, release all previous scores received by a test subject on a test to anyone designated by the test subject to receive the current score.

This section shall not be construed to prohibit release of scores and other information in the possession of a test agency for purposes of research leading to studies and reports concerning the tests themselves. Such studies and reports must contain no information identifiable with any individual test subject.

§ 345. Regulations. The commissioner shall promulgate regulations to implement the provisions of this article.

§ 346. Violations. Any test agency which violates any section of this article shall be liable for a civil penalty of not more than five hundred dollars for each violation.

§ 347. Severability. If any provision of this article shall be declared unconstitutional or invalid, the other provisions shall remain in effect notwithstanding.

§ 2. This act shall take effect on the first day of January next succeeding the date on which it shall have become a law.

Attachment

THE UNIVERSITY OF THE STATE OF NEW YORK
THE STATE EDUCATION DEPARTMENT
ALBANY, NEW YORK 12234

New York State Standardized Testing Law

(Education Law Article 7-A as enacted by Chapter 672 of the Laws of 1979)

QUESTIONS AND ANSWERS
October 1, 1979

Below are some commonly asked questions concerning the Standardized Testing Law. The answers reflect the opinion of the State Education Department.

Section 340

1. **Question:** To what tests does the Law apply?

Answer: The Law applies to any test which is given at the expense of the test subject and is designed for use or in fact is used in the process of selection for postsecondary or professional school admissions, whether the test is administered in New York State or the results of the test are provided to institutions located within the State.

2. **Question:** The definition of "test agency" contained in the Law includes individuals who administer a test. Does this mean that proctors and organizations which simply provide testing space are obligated to furnish test information?

Answer: No. One who simply provides space or proctoring services is not considered a test agency. Primary responsibility for compliance with the Law rests with the sponsor. Specific responsibilities may be delegated by the sponsor to the test publisher or another agency.

3. **Question:** Does the Law apply only to machine scorable, multiple-choice tests?

Answer: No, the Law also applies to any form of test which meets the definition in Section 340.

4. **Question:** Is a test which is in fact used in the admissions process subject to the provisions of the Law even if it was not designed or validated for use in the admissions process?

Answer: Yes. The Law specifically applies to tests designed for or used in the admissions process.

5. **Question:** Does the fact that a test score is included in an applicant's admission folder automatically subject the test to the Law? For example a high school record submitted to a college might contain test results not required by the postsecondary school.

Answer: No. The Law applies only to tests required and used in the admissions process.

6. **Question:** Is a test administered and used outside the State subject to the Law if it was taken by a New York State resident?

Answer: No. The Law would not apply simply because the test was taken by a resident of the State or simply because scores were mailed to a candidate living in the State.

7. **Question:** Does the phrase "...at the expense of the test subject..." include tests for which fees are paid indirectly by the test subject?

Answer: Yes. The Law does not differentiate between direct fees paid to the test publisher and indirect fees which may be included in an application fee required by the school.

Section 341

8. **Question:** Are test agencies obligated to release reports and data prepared for individual institutions?

Answer: Yes. The Law does not exempt any type of report or data. However, information identifiable with an individual test subject must be deleted.

9. **Question:** Are test agencies obligated to obtain and file studies prepared by others?

Answer: Under certain circumstances, yes. Test agencies are required to file those reports which they prepare or cause to have prepared, or for which they provide the data. Test agencies are not required to file studies prepared by others which are based solely on data available to the general public unless such studies are undertaken at the request of the test agency.

Section 342

10. **Question:** Are test agencies required to return the original answer paper to the test subject?

Answer: No, a legible copy is sufficient. However, a computer generated list of the test subject's answers would not be acceptable.

11. **Question:** If a test agency furnishes the raw score with the original standard score report, is the agency obligated to provide the subject's raw score a second time if asked to do so?

Answer: No.

12. **Question:** Subdivision 2 of Section 342, which requires that certain information concerning a given test be sent to each test subject upon request, provides a 90 day deadline for the making of such requests. Does that deadline apply only to requests for the subject's raw score (paragraph c), or does it apply as well to requests for copies of the test questions (paragraph a) and the subject's answer sheet and the correct answer sheet (paragraph b)?

Answer: The 90 day deadline for making requests applies to all of the information specified in paragraphs a, b and c.

13. **Question:** Do "direct cost" include the salaries of extra staff needed to comply with the Law?

Answer: Yes, direct costs are those additional costs which a testing agency will incur in meeting the mandates of the Law.

14. **Question:** Must the test agency provide a test subject with the information specified in Section 342.2 even if the test subject does not appear for testing or orders that the test not be scored?

Answer: No

Section 343

15. **Question:** Must the "statements" mentioned in Section 343.1.c refer to the specific form of the test to be administered?

Answer: No. The statements may be based on information collected from other forms of the test.

16. Question: Subdivision c of Section 343 requires the filing of "available correlations". Are test agencies obligated to prepare and provide correlations between test scores and other factors if such correlations would not be prepared except for the provisions of the Law?

Answer: No. Test agencies are obligated to provide correlations only if prepared for other purposes.

Section 344

17. Question: Section 344 provides that the score of any test subject shall not be released to others unless "specifically" authorized by the test subject. Does that mean that a separate and specific request must be made by the test subject for each score release?

Answer: No. The test subject must specifically authorize the release of scores, but this authorization would permit the release of scores to all institutions which request them, unless limited by the subject to specific institutions.

Attachment 3

THE UNIVERSITY OF THE STATE OF NEW YORK
THE STATE EDUCATION DEPARTMENT
ALBANY, NEW YORK 12234

AMENDMENT TO REGULATIONS OF THE COMMISSIONER OF EDUCATION

Pursuant to Section 345 of the Education Law

The Regulations of the Commissioner of Education
are amended, effective January 1, 1980,
by the addition of a new part 56 to read as follows:

Section 56.1 Definitions.

Sponsor as used in this part shall include any organization, association, corporation, partnership or individual responsible for creation of a test.

Section 56.2 Responsibility for compliance.

Compliance with the requirements of Education Law Article 7-A and of these regulations shall be the responsibility of the sponsor. A sponsor which delegates specific responsibilities to the test publisher or another agent shall so advise the Department in writing within 15 days after making such delegation.

Section 56.3 Filing of test information and materials.

(a) On or before January 15, 1980 and the fifteenth day of each subsequent testing year used by the sponsor, the sponsor shall file with the Commissioner of Education:

(1) A list of the names of the tests of such sponsor which are subject to the provisions of Education Law Article 7-A, and

(2) The beginning and ending dates of the test year used by the sponsor, and

(3) A calendar of the regularly scheduled administration dates and anticipated score release dates for such administrations during the test year, and

(4) A copy of each general publication prepared for routine distribution to test subjects, and

(5) A list of the specific duties required for compliance with the law which the sponsor wishes to delegate to another agent. The names and addresses of the agent or agents should be specified.

(b) Background reports and statistical data required by Education Law Section 341 to be filed with the Commissioner of Education shall be filed within thirty days after completion.

Section 56.4 Transmittal of information to test subjects.

(a) Each request by a test subject for the material described in Education Law Section 342, Subdivision 2 shall be in writing and signed by the test subject. Such material shall be transmitted to the test subject within 30 days after receipt of each such request by the test sponsor or the sponsor's designated agent.

(b) Each test subject's answer sheet shall be considered to include any documents or materials prepared or produced by the test subject during the test administration, provided such documents or products are evaluated by the test sponsor or the sponsor's designated agent for purposes of determining the subject's score or performance on the test.

(c) The test sponsor or the sponsor's designated agent may provide the test subject with a photocopy of the test subject's answer sheet.

Postsecondary and Professional School Admission Testing Programs Identified by Sponsors or Publishers

Grouped by Tentative Plans of Sponsors

September 20, 1979

Programs to be continued with no reduction in services

None

Programs to be continued with some reduction in services

ACT Assessment Program
Graduate Management Admissions Test
Graduate Record Examination
Law School Admissions Test
Preliminary Scholastic Aptitude Test/National Merit Scholarship Qualifying Test*
Scholastic Aptitude Test and Achievement Tests

Programs to be withdrawn from use in New York State

Allied Health Entrance Examination
Allied Health Professions Admission Test
Aptitude for Practical Nursing Examination
Aptitude Test for Allied Health Programs
Dental Admission Test
Doppelt Mathematical Reasoning Test
Entrance Examination for Schools of Health-Related Technologies
Entrance Examination for Schools of Nursing
Entrance Examination for Schools of Practical/Vocational Nursing
Health Occupations Aptitude Examination
Medical College Admissions Test
Miller Analogies Test
Minnesota Engineering Analogies Test
Nursing School Aptitude Examination
Optometry College Admissions Test
Pharmacy College Admissions Test
Pre-Admission Assessment for Practical Nursing
Pre-Nursing and Guidance Examination
Pruebas de Aptitud Academica and Pruebas de Aprovechamiento Academico
Veterinary Aptitude Test

---

*The PSAT/NMSQT is specifically named in the law. According to the publisher, however, it is not used for admission purposes.

FEDERATION OF ASSOCIATIONS
OF SCHOOLS OF THE HEALTH PROFESSIONS
One Dupont Circle, Suite 810 Washington, D.C. 20036

OFFICE OF CHAIRMAN

STATEMENT OF THE
FEDERATION OF ASSOCIATIONS OF SCHOOLS OF THE HEALTH PROFESSIONS
ON PROPOSED TESTING LEGISLATION
(H.R. 3564 and H.R. 4949)

Submitted to the House Committee on Education and Labor

Subcommittee on Elementary, Secondary and Vocational Education

On September 25, 1979

Mr. Chairman:

As educators in health professions fields, we uphold the institution of standardized testing. While they should not be the sole factor in evaluating a student, test results, combined with other academic and sociological data, help to give the complete picture of the student. The demise of such testing would negate many benefits that our schools -- and all higher education -- have been realizing over the years.

Fair treatment and equal rights for all examinees, we believe, are essential. We share the concerns expressed by many that the testing procedure be a fair one, that individuals being examined be fully informed about the tests' nature and use, and that some mechanism be in place to assure the examinees that test scores accurately report performance on the examination.

Nationally endorsed standards and safeguards regarding standardized test development and administration have been in place for a number of years. These standards, developed by testing and measurement specialists through their professional associations, have fostered developments of reliable test instruments as well as sought to protect the rights of examinees. This program is dynamic, continually under review and revision to assure that standardized testing programs remain responsive to these standards. This historical precedent has convinced us that such a program is the appropriate mechanism to continue to assure that the rights of examinees be protected. Thus, at the present time, it has not been demonstrated that government regulation in this area is necessary.

We would strongly oppose a new law that would allow public disclosure of test questions and answers, since this would destroy many of the important advantages of standardized testing.

....Continued

MEMBER ORGANIZATIONS: American Association of Colleges of Osteopathic Medicine • American Association of Colleges of Pharmacy • American Association of Colleges of Podiatric Medicine • American Society of Allied Health Professions • Association of American Medical Colleges • Association of American Veterinary Medical Colleges • Association of Schools and Colleges of Optometry • Association of Schools of Public Health • Association of University Programs in Health Administration • National League for Nursing (Council on Baccalaureate and Higher Degree Programs)

A major raison d'être of standardized tests is that they provide a standard measure through time and thus provide a mechanism to evaluate individuals, programs, and indeed whole educational systems. If a new examination had to be written for each test administration, as would be the case if public disclosure of the test items existed, these advantages of standardized testing would be lost. Such a requirement would also make it impractical for the companies to improve tests by eliminating many of the statistical procedures available when the same test questions are used over time. This would decrease the power to differentiate those students who have knowledge or ability from those who do not; it would decrease the high reliability which characterizes most national standardized tests.

Another problem with disclosing test items arises in connection with the more specialized academic areas in which the writing of an unlimited number of test questions is constrained by the more narrowed field. It is likely that within a relatively short period of time nearly all the most desirable test questions would have been given, thus providing the advantage of knowing the questions in advance to future examinees.

Cost is another factor. In the health field the number of examinees is generally small. Any increase in the costs of test development and administration must be spread among a small audience, representing for some students an additional artificial, nonacademic barrier to admission.

We would like to point out two paradoxical, and perhaps unforeseen, effects of proposed testing legislation. First, the unrealistically high price tags that would result for tests given to small numbers of examinees would cause many testing firms to discontinue their testing programs. Only the largest firms administering tests to large numbers of students would be able to keep their examination fees reasonable enough to continue their testing programs. This, unfortunately, would have a tendency to create a monopoly situation.

Second, without the aid of standardized tests, student admissions decisions would be based upon more subjective information. Contrary to the belief of some, standardized test scores, when properly used, actually have afforded institutions the opportunity to diversify their student bodies. These objective and standard measures of students from all educational institutions have allowed admissions committees to evaluate candidates independent of biases and suspicions of grades earned at known and unknown schools. Also, when used as a diagnostic indicator of student abilities, standardized tests have allowed institutions to develop academic support programs for groups of students desired by the institution yet previously disqualified because of records of poor performance.

We appreciate the opportunity to present this statement, which was unanimously endorsed by the current membership of the Federation of Associations of Schools of the Health Professions: American Association of Colleges of Osteopathic Medicine, American Association of Colleges of Pharmacy, American Association of Colleges of Podiatric Medicine, American Society of Allied Health Professions, Association of American Medical Colleges, Association of American Veterinary Medical Colleges, Associations of Schools and Colleges of Optometry, Association of Schools of Public Health, Association of University Programs in Health Administration, National League for Nursing (Council on Baccalaureate and Higher Degree Programs).

PREPARED STATEMENT OF DR. CLIFF SJOGREN, DIRECTOR OF ADMISSIONS, UNIVERSITY OF MICHIGAN—ANN ARBOR

I submit this statement for the record first as the Director of Admissions at the University of Michigan, a large institution of higher learning that consists of eight undergraduate schools and colleges and ten graduate and professional schools, each of which employs selective and competitive admissions.

I also write on behalf of the American Association of Collegiate Registrars and Admissions Officers (AACRAO), a national, non-profit educational association which is composed of approximately 2,000 accredited colleges and universities that are represented by over 6,900 individual members from the field of recruitment, admissions, registration and records, student financial aid, and institutional research at the undergraduate, graduate, and professional levels. -

My colleagues and I share with the sponsors of HR 4949 and HR 3564 their basic philosophy of providing equal and the best possible higher education opportunities for all of the youth of our nation--and I might add we support these opportunities for all adults as well. We do not share with the sponsors and the proponents of this proposed legislation the thesis that tests are the sole or primary criterion for admissions to higher education. Nor do we believe that the enactment of Federal legislation which will cause nationally recognized and respected testing agencies to divulge test booklets, answers, and associated data to examinees or to private researchers will give greater or easier access to higher education for those who seek admissions.

The professional and graduate schools have already voiced their concerns before this committee on problems they will encounter if the bills are passed. We support them and their positions on handicaps that will be imposed on those schools with the pending disclosure requirements.

The undergraduate recruitment and admissions process and the outreach programs well established through the nation in higher education to embrace the economically and educationally disadvantaged and the minorities are, we believe, in equal jeopardy. The possibility of expanding educational opportunities will also be drastically reduced if such a law is enacted.

The concerns I wish to place before the committee in this statement are limited to the possible and probable adverse impact such a law would have on the undergraduate admisisons process.

To elucidate: At the University of Michigan the schools and colleges are selective as they enroll only students for whom a high probability of success can be predicted. Because all qualified students can not be admitted due to space limitations, admissions are also competitive.

Admissions to the University is based on a complex and interrelated set of factors that combine to provide evidence of academic potential and educational readiness. The secondary school experience, which includes grades, class rank, quality of courses elected, and the success pattern of previous graduates at the University, provide the most significant influence on admissions decisions. The Scholastic Aptitude Test (SAT) is also an important factor in the decision of several candidates. No single criterion will guarantee either admission or rejection. The importance of the test score diminishes, however, as consistency and quality are revealed in the secondary school record. For example, an applicant to Michigan who attended an accredited high school, earned a B+ average and had completed a rigorous program of studies will be admitted regardless of SAT performance. Of further significance is the example of an applicant, frequently a minority or rural student, who attended a school of marginal quality and achieved a C+ average, but who presented strong SAT's, or a commendable standard on a biology achievement test would be given a close look since there is evidence that he/she is capable of learning even though his/her educational environment might have retarded scholastic achievement.

The best predictor of academic behavior is previous academic behavior. Reliable test data, therefore, gives the applicant the opportunity to demonstrate competence even though his/her previous record was mediocre. The University, as is true at nearly all, if not all institutions offering undergraduate courses, does not employ "cut-off" scores nor do test scores prevent an otherwise well-qualified candidate from being admitted. Test scores supplement the more important factors of classroom performance. High test scores generally have more relevance in the admissions decision than do low test scores.

Reliable test data help hold back the negative influence of grade inflation. It would be impossible for Michigan to assess the quality of the estimated 1,000 secondary schools from which applications are received. Standardized measurements are needed to maintain our efforts to provide fair treatment to all candidates for admissions and to not rely solely on subjective teacher evaluation. The admissions process at Michigan, therefore, depends on reliable and secure standardized examinations that are frequently administered at low costs to the students. We feel that HR 4949 greatly reduces our efforts to service all applicants fairly.

Previous testimony on HR 4949 suggests that tests are particularly descriminatory against minority and low income students. At Michigan, the opposite is true. As previously stated, reliable test data will assist us in identifying students with academic promise. Further, as we attempt to increase the diversity of students represented on campus, we seek to find the best students possible from a variety of groups of people with similar characteristics. In a sense, these students compete within their own groups as we seek to enroll the most qualified students from each of several groups that include nonresidents, rural students, foreign students, black students, Michigan residents, native Americans, and others. Until a satisfactory representation has been achieved for each group, we extend admissions priorities for that group. Test scores are an important means of identifying the most capable candidates from each of the populations.

I would now offer some general comments on the proposed legislation. Having read carefully the statements of testimony and having completed nineteen years as an admissions officer, served periodically as a consultant to several prestigious institution of higher learning and served five years as a secondary school teacher and counselor, I bring the following observations to the attention of the committee:

1. At the undergraduate level, most institutions use multiple selection factors for admissions. The position is fully supported by valid research released this month in a survey of our AACRAO membership which revealed that only 1.7% of responding institutions reported that test scores were the single most important factor in admissions decisions. This does take from tests their significance as a singular important factor and disputes their use for discrimination.

2. The propose legislation will increase test costs to students, reduce the number of test administrations, and reduce the number of test sites, all of which will work to the disadvantage of the student, particularly those who are poor and who reside in geographically isolated areas.

3. A single test item that has been disclosed in the past and is used on a new form damages the validity of the test and will give special advantages to those who enroll in profit-making test coaching schools.

4. College and university personnel today are much more knowledgeable and sophisticated in the use of tests for admissions purposes. The improvement in this respect since the 1960's and early 1970's has been remarkable.

5. Reliable standardized tests more often work for, rather than against, an applicant who has academic potential.

6. Because "cut-off" scores are not and should not be used, uncontaminated research on the degree to which tests predict academic success would be impossible. Indeed, tests should be used only to supplement the more reliable secondary school achievement record. Consequently, it would be impossible to indicate the percentage of improvement tests provide in predicting academic success as required by HR 4949.

7. Minority, rural, and low income students who have academic potential can often be identified by reliable standardized tests where they might be missed by evaluating school performances only. Test agencies now compile data at the request of an institution to enable the institution to reach these students. Institutions will tend to drop this type of program rather than risk the chance of

having their data now supplied in confidence to testing agencies released to the public and taken out of context to rank them against other institutions, for example. The disclosure provisions in the proposed law would make this possible under the Freedom of Information Act. Many institutions would hesitate to provide information on students tests scores for fear that personally identifiable data might be possible to ascertain.

8. The requirement of the pending legislation for documentation of the ability of tests to predict career success is questionable. While the bill places the onus on the testing agencies, the basic data would have to come from the institutions. There is no available evidence that tests can predict career success, if, in fact, we can adequately define what "career success" is.

9. The proposed legislation, if enacted into law, leads to a dangerous and real possibility of excessive Federal controls of the admissions process.

10. Test agencies (The College Board, the American College Test, etc.) have substantially improved and are continuing to improve their literature and other communicative devices on the proper uses and limitations of their tests. Professional associations (AACRAO, ACE, NACAC and others) have also monitored the activities of their memberships and "statements of good practices" have been written and widely circulated.

The concerns I have cited are not just at the University of Michigan. They are the concerns of admissions officers throughout the United States. I submit that the major undergraduate admissions programs now offered by the colleges and universities of this Nation have been designed and have been proven successful in reaching out to many more of the youth of America---the future leaders of our Nation--than was previously thought possible. Great strides have been made, with the help of the testing agencies, to encourage many young persons who had previously been without hope of a college education to participate. A look at a college or university campus today will show you a mix of America hardly dreamed of even two decades ago. To destroy, through the enactment of Federal legislation, a system known to be effective without a tested alternate system in place will weaken the affirmative action programs considerably. A trial and error period to replace the present system will deprive many of our capable youth of a college education unnecessarily. It will cause a hardship to many citizens and to institutions of higher learning who would like to help them.

In summary, I urge the committee to either cease further consideration of both HR 4949 and HR 3564 or delay any further action until the full effects of the New York law can be properly evaluated. If the bill is enacted into law, the fair and educationally sound processes of admissions as now practiced will needlessly be changed. Institutions and test agencies are eagerly and effectively meeting their responsibilities to their constituents. We would invite a closer examination of our practices before serious damage occurs.

PREPARED STATEMENT OF THE NATIONAL BOARD OF MEDICAL EXAMINERS, PHILADELPHIA, PA.

The purpose of this document is to comment on the legislation concerning educational and occupational admissions testing (H.R. 3564 and H.R. 4949) now before the Subcommittee on Elementary, Secondary, and Vocational Education of the U.S. House of Representatives. The National Board of Medical Examiners views these two bills with concern because they contain provisions which pose serious threats to the quality of educational and occupational testing.

The National Board of Medical Examiners is a private, non-profit organization which, since 1915, has developed examinations leading to licensure and certification in medicine and more recently in certain allied health professions. These examinations are accepted by state regulatory agencies which are responsible for the licensure and certification of physicians and allied health personnel. As such, they provide the only external evaluation of professional qualifications for practice that is independent of both the educational institutions and the licensing authorities. Successful performance on the required qualifying examination results in licensure or certification and thereby in legal authorization to practice the healing arts in this country. Because consistent and substantial improvements in these examinations have been achieved over the past 65 years, the public has come to expect that satisfactory test performance indicates that a candidate for professional practice possesses the knowledge and problem-solving skills required for safe and effective health care delivery.

With respect to occupational testing in the professions, some provisions of H.R. 3564 and H.R. 4949 jeopardize the capacity of such testing programs to evaluate accurately the qualifications of candidates for professional licensure and certification. These provisions thereby weaken the degree to which these testing programs can protect the public welfare, and the National Board feels it must oppose such provisions in order to fulfill its responsibility to the public.

In addition to its responsibility to the public welfare, the National Board of Medical Examiners also recognizes its responsibility to those who take its examinations to make those examinations fair and valid tests of relevant professional and scientific knowledge and skills. In fulfilling its responsibility to examinees, the National Board provides detailed descriptions of its examinations which are revised annually ("Bulletin of Information and Description of National Board Examinations"), and a reference book is available which describes all of the operational procedures and research and development efforts of the National Board of Medical Examiners (Hubbard, John P., "Measuring Medical Education; The Tests and the Experience of the National Board of Medical Examiners." Second Edition, 1978.) These two publications provide detailed information to examinees concerning the subjects covered in each National Board examination, the types of professional capabilities that are tested, the scoring and standard setting procedures employed, the standards required to pass each examination, the National Board's policy regarding the reporting of examination scores, and the rights of examinees to determine which agencies will receive these scores. Examinees are also given information that is helpful in interpreting their scores on the individual subjects covered by each examination. This enables examinees to identify areas of relative strength and weakness. In addition, examinees are advised of their rights to have their examination scores rechecked by hand. Samples of the content and format of test questions drawn from previous administrations of each examination are also provided. Statistical analyses and validation studies of National Board examinations are also described, and reprints are available at no charge on request.

These materials provide all examinees with a comprehensive description of each examination, a thorough orientation to the testing materials they will encounter, and a full disclosure of their rights to privacy and due process. Such materials have been available to those who take National Board examinations for many years. The National Board of Medical Examiners is aware that similar materials are available in other testing programs and constitute a widely followed standard of practice within the measurement profession.

However, the reader of H.R. 3564 and H.R. 4949 would conclude from the provisions of these two bills that little or no such information is provided to examinees. Although the National Board supports the disclosure of information to examinees, it is of the opinion that little can be achieved by passing these two bills when they mandate the disclosure of information which, by and large, is already provided routinely by test sponsors and testing agencies. The National Board does not believe that the sponsors of these two bills have demonstrated that the provisions in them will achieve added enhancements in the quality of occupational testing which have not already been achieved through the voluntary efforts of the professions and of testing agencies.

Indeed, it is the provisions of these bills which are not now standards of practice within the measurement profession that seriously concern the National Board of Medical Examiners. Rather than advance the quality of occupational testing, these provisions would seriously endanger its effectiveness.

The first of these provisions is contained in H.R. 3564 (Section 6(c)) and prohibits the establishment of passing scores "on the basis of the relative distribution of scores of other test subjects." Setting examination standards is an extremely complex activity requiring the integration of accepted psychometric principles and techniques with the careful judgment of multiple experts in a profession. In examinations that are used to admit individuals to a profession, the examination standard must be set at a level which maximizes the likelihood that individual examinees have been correctly classified as either qualified or unqualified for professional practice. At the present time, there are two basic approaches to setting examination standards. One is to set a relative standard based on the performance of a reference group of examinees. This approach would be prohibited by H.R. 3564. The other approach is to set an absolute standard without considering the performance of any group of examinees and consequently, without considering any objective measure of the difficulty of the questions included in a given examination. The relative standard can be described as one in which a fixed percentage of examinees in some group must fail; the absolute standard can be described as one in which a fixed proportion of questions must be answered correctly in order to pass a given examination regardless of the actual difficulty of that examination. Both approaches are subjective; both require an element of judgment; and neither has been shown in the scientific literature to be superior to the other. However, both approaches can be implemented in a fair and logical manner such that once an examination standard has been established, it can be applied uniformly from one test administration to the next. The provision in H.R. 3564 (Section 6(c)) which prohibits the use of one standard setting approach when there is no sound evidence in the scientific literature or even consensus within the measurement profession regarding the superiority of either approach is an arbitrary and unwarranted restriction which in no way enhances the effectiveness or the fairness of standard setting. Therefore, the National Board of Medical Examiners strongly urges the elimination of Section 6(c) from H.R. 3564.

Another provision that is particularly troubling to the National Board of Medical Examiners is the requirement in H.R. 4949 (Section 5(a)(1)(A-B)) for the disclosure of test questions and answers following each administration of an examination. Licensure and certification examinations in the health professions are designed to test those aspects of knowledge and problem-solving skills that are essential requisites to safe and effective health care delivery. The three examinations which comprise the National Board Examinations are developed each year through the major commitment of time and effort of over 100 experts in the basic and clinical sciences. It is conservatively estimated that these experts devote more than 4,000 man hours to the development of new test questions for National Board Examinations. Of the total number of items reviewed each year, only fifty per cent are judged to meet the high standards of scientific correctness and psychometric quality required of examinations that will be used to assess individual qualifications for medical licensure. The task of developing high quality test items is a complex, arduous, and time-consuming one. Once test items hav been administered in actual National Board Examinations and are found to be of sufficient scientific and psychometric quality, they are classified according to a multi-dimensional categorization system and stored in a computer-managed test item library for use on future National Board Examinations. This process is a significant factor in maintaining the high quality of these examination programs because it provides a mechanism whereby previously tested items can be reused on subsequent examinations, thus ensuring not only the quality but the equivalence of examinations from year to year. Multiple administrations of high quality test items not only ensures the continued validity and equivalence of examinations from year to year, but also results in cost savings which help keep the fees for these examination programs within the financial means of the students who must take them.

The requirement to disclose all test questions and answers following each administration of an examination will impair the effective reuse of high quality test items and in so doing will have a long-term negative effect on the quality of test materials found in these examination programs. Moreover, this provision will increase by a significant proportion the costs of test development, test administration, test production, and scoring and analysis. Because test items that are used in calculating an examinee's raw score must be disclosed following the administration of the test, these items will not be reusable on subsequent examinations. This will have two consequences. First, it will result in a thirty to forty per cent increase in the number of new test questions that must be developed each year. Moreover, in order to include a sufficient number of test items for the establishment of test equivalence from year to year, the length of each examination will have to be increased by approximately twenty-five per cent. Students who would normally sit for a two-day examination will now have to be tested for two and a half days. Costs for developing test items, renting examination centers, paying fees to proctors, printing, scoring and analyzing such a test will increase substantially. Because most occupational testing programs involve a much smaller number of examinees than is found in admissions testing programs, the impact of increased costs on individual examinees is much greater in occupational testing programs.

Another serious consequence of test item disclosure is that it will destroy the security of the National Board's test item pool which now forms the basis of constructing three other examinations used to assess qualifications for graduate medical training and medical licensure. Specifically, the licensure examination of the Federation of State Medical Boards which is accepted by all licensing authorities in the United States and its territories could no longer be developed as it is today from previously used and hence statistically calibrated test items. The Visa Qualifying Examination and the examination of the Educational Commission on Foreign Medical Graduates would be similarly affected. These examinations are used to evaluate the professional qualifications of foreign-trained physicians who are seeking entry into the U.S. educational and health care delivery system. Under the test item disclosure provision, these examinations could no longer be developed by drawing upon the same test items that have been used to evaluate the professional qualifications of U.S. medical students. During the past decade, an enormous effort has been made, and supported in part by the actions of Congress, to ensure that all physicians in residency training and all candidates for medical licensure in this country meet the same standards for knowledge and problem-solving skills. The disclosure of test items will severely compromise the National Board of Medical Examiners' capacity to meet this national goal and ensure the content and psychometric equivalence of these testing programs.

Potentially the single most serious consequence of the test item disclosure requirement will be its negative impact on learning. Without doubt, the sudden availability of thousands of test questions and corresponding answers will divert the attention of medical students and students in other health professions away from the scholarly study of the biomedical sciences. Instead, students will be strongly tempted to engage in crash cramming, relying on disclosed test material from previous administrations. Wholesale review of test questions and answers will result in a superficial and only transitory acquisition of knowledge. A deeper understanding of scientific facts and principles will be lost, and will instead be replaced by a superficial and glib awareness of the correct answers to test questions. The only way in which professional competence is acquired is through conscientious study and ongoing experience in the application of scientific principles. Any activity which distracts a professional student from that basic mission seriously weakens professional growth and development and produces poorly prepared individuals who lack the firm understanding, skill, and sound judgment required for safe professional practice.

The descriptive materials routinely available to examinees already provide adequate and effective orientation and guides to the preparation for examinations. There can be no further benefits to be derived from disclosing test questions and answers following each administration of an examination. Indeed, such a practice will jeopardize the measurement of accuracy of such examinations and will interfere with the thorough and sound acquisition of scientific knowledge and understanding. The proponents of test item disclosure have failed to document the supposed gains to be derived from such a requirement. In fact, no such gains exist. Instead, only serious threats to the valid assessment of professional competence can be expected. It has been observed that "to the degree that we are able to measure competence, to just that same degree will medical competence be available to the people." (Womack, N.A. "The National Board of Medical Examiners - A Perspective." The National Board Examiner, 1963, 10(7): 1-4.) The National Board of Medical Examiners urges that H.R. 3564 and H.R. 4949 not be approved.

PREPARED STATEMENT OF THE AMERICAN SOCIETY OF CLINICAL PATHOLOGISTS

The American Society of Clinical Pathologists (ASCP) is a non-profit, educational and scientific medical specialty society representing nearly 23,000 pathologists and other medical laboratory professionals. Our members practice in a wide variety of laboratory environments: hospitals, universities, independent laboratories, military and veterans' hospitals, and federal, state and local government facilities. Our organization is vitally concerned with the quality of health care in this nation, and we are committed to upholding the very highest standards in the clinical laboratory. The two major activities of the ASCP are the development of educational programs and products for laboratory professionals, and the certification of non-physician medical laboratory personnel. It is because of this latter activity that we are concerned with legislation presently under consideration by this Subcommittee, and we appreciate the opportunity to submit comments on H.R. 3564 and H.R. 4949.

The ASCP and its Board of Registry have been involved in the certification of medical laboratory personnel since 1928. To date, the Board of Registry has certified just under 200,000 individuals. At present, certification is offered in three generalist categories (Medical Technologist, Medical Laboratory Technician and Clinical Laboratory Assistant) and in the following specialty areas: Microbiology, Hematology, Chemistry, Blood Banking, Immunology, Nuclear Medicine, Histotechnology and Cytotechnology. The Board of Registry is a private, voluntary certification agency. Examinations are administered twice a year throughout the country. The Board of Registry itself is made up of six pathologists, six medical technologists, two public members, and one representative from each of the following organizations: American Academy of Microbiology, American Association of Blood Banks, American Society of Cytology, National Registry in Clinical Chemistry, American Society of Hematology, and the National Society for Histotechnology. The present Chairperson of the Board is an ASCP-registered medical technologist. This Board and its examination committees -- approximately 100 professionals -- are responsible for the development and updating of the various examination forms and the pool of test items. They have staff support of 36 individuals in the ASCP Board of Registry offices in Chicago. This staff includes testing and education professionals.

Our concern with H.R. 3564 and H.R. 4949 grows out of our present involvement with and fifty-one years experience in testing and certification. Although it is our understanding that, as they are written, neither of these bills would apply to private, voluntary certifying organizations such as ours, we feel that any agency involved in or interested in testing should be closely monitoring and participating in the consideration of legislation such as this. We must state our strong opposition to these two bills.

As members of the Subcommittee may know, in the health care field, private certifying agencies have, with Federal support, recently established the National Commission for Health Certifying Agencies. This organization will set standards, evaluate, and accredit the various agencies which certify health care professionals. It is, we believe, a fine example of what can be achieved in the private sector, with government support and cooperation, rather than by government regulation. It is a model which this Subcommittee might study as an alternative to the proposed legislation.

Though the goals of the sponsors of H.R. 3564 and H.R. 4949 are laudable, it seems apparent that these bills would not achieve these goals. This fact has been pointed out by both witnesses and members of the Subcommittee during recent hearings on this legislation. Based on our experience in testing, our close reading of the bills, and our observation of the problems already being encountered in the state of New York where similar legislation has been passed, we certainly concur with that position. Indeed, we believe that the implementation of this legislation might prove harmful to the very individuals its sponsors wish to help.

Further, there appears to be no established need for many of the provisions in this legislation, particularly those of H.R. 3564. Applicants for examinations are presently provided with a considerable amount of information prior to the examination, and those who sit for the examination are given explanatory material with their test results. This is certainly true of the Board of Registry examination, and we have included, as attachments to this statement, samples of the material our examinees receive, prior to and after the examination. This material includes a subject outline of information upon which they will be tested, samples of the kinds of questions to appear on the examination, test scores broken down by subject category, and a list of individuals within the Board of Registry staff to contact should they have need of further information. It is our understanding that many testing agencies have made available to this Subcommittee similar samples of the types of material which they provide to examinees. H.R. 3564 and parts of H.R. 4949 would simply be a codification of practices presently in effect in testing agencies.

In testimony before this Subcommittee important points have been made which refute the claims made by proponents of these bills regarding secrecy of examinations. Since most examinations are in fact developed, screened and evaluated by professionals in the subject areas being tested, it can hardly be held that professionals in the field do not have access to the examination. The Medical College Admissions Test is developed and evaluated by physicians and physician-educators, the Law School Admission Test by attorneys and law school faculty members. The ASCP Board of Registry examination is developed and evaluated by laboratory professionals -- pathologists, medical technologists and clinical scientists.

Our major concern with this legislation is the provision (Section 5 of H.R. 4949) which would require disclosure of the examination itself and correct answers to the test questions. It is our belief that this provision, if enacted, would be truly destructive to many examinations presently administered. Though H.R. 4949 in its present form applies only to educational admissions examinations, we are concerned about possible amendments which would combine it with H.R. 3564, or otherwise extend its provisions to occupational examinations. There are examples of tests within both categories of examination -- educational and occupational -- which are limited because they are designed to test certain types of knowledge, or because they take a unique approach to the knowledge for which they test. Disclosure of the examination questions and answers, so that those particular questions could not be used again, would have serious consequences for these achievement -- as opposed to aptitude -- examinations.

In our Board of Registry examination, for example, the approach taken in designing a test item is as important as the substance of the material being tested. The examination is based on a defined content, appropriate methods of construction, criterion validity and reliability. Because we are testing for acquisition of very specialized knowledge, because we seek to assess more than just cognition, our test items are designed accordingly. Further, because of developments in health-care technology, and the ever-changing nature of the practice of medicine, you can well imagine the dynamic nature of the base of knowledge on which examinees are being tested. It differs somewhat from the material used in an admissions examination covering general knowledge of, for example, mathematics, grammar or history.

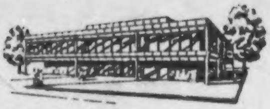
We certainly agree with the sponsors of this legislation that a feedback mechanism is needed, so that an examinee can learn from the examination process. It is for this reason that we break down our test results into subject categories which correspond to the categories in the subject outline given to the applicant prior to the examination. (see attachments) The examinee is told how many questions he missed in each of the categories of the subject outline. Told that he missed 8 out of 29 questions in bacteriology, for example, he will go back to that subject area and review it to determine where his mistakes were. If, on the other hand, he had questions and answers, he would check to see what the correct answer was and go no further. The efforts of test developers would be better spent in constructing more sophisticated methods of providing meaningful feedback to examinees than in developing more and more test items to replace those disclosed after each test administration.

We estimate that our overall cost for development of one test item is $300-$400, and this figure does not include a factor for the many hours of time donated by volunteer members of our Society. The members of our Board of Registry and its examination committees -- all of them laboratory professionals, scientists, educators -- meet several times during the year to work on test development and updating and also put in many hours of time working individually on test items. They receive no payment for this work.

We are able to maintain the very highest standards in our examination because of the constant efforts of these professionals and of our Board of Registry staff to develop and update our test item pool. If each of our examination forms were made public, with correct answers, that item pool would soon be depleted, our costs for development would increase, and the process by which we certify laboratory professionals would be altered significantly. We believe that many testing agencies would find themselves in a similar situation were this legislation to be enacted.

Based on our experience in the testing field and our knowledge of test development and administration, we urge the Subcommittee to re-examine these bills, to compare the legislation itself to the goals its sponsors seek to achieve, and to give careful consideration to the testimony of testing professionals and educators. This legislation does not serve those who are now taking examinations or those who will sit for examinations of all types in the future, nor is it in the best interests of education or occupational certification. We strongly oppose H.R. 3564 and H.R. 4949.

Thank you for this opportunity to submit our comments.



AMERICAN SOCIETY OF CLINICAL PATHOLOGISTS
2100 WEST HARRISON STREET • CHICAGO, ILLINOIS 60612 • (312) 738-1336

BOARD OF REGISTRY, JOHN W. GREEN, JR., M.D.

MEMORANDUM

TO: Applicants for Board of Registry Examinations

FROM: Office of the Director

SUBJECT: Examination Procedures

The enclosed documents have been prepared to explain important Board of Registry examination procedures and to provide you with a general content outline for the specific examination for which you have applied. With regard to this outline, it should be noted that specific references are *not* mentioned. The Board of Registry feels that since the laboratory sciences literature is so voluminous, it is inappropriate to officially review, recommend, or endorse any material in the form of reading lists or study guides. The Board can suggest that an applicant use any standard texts or references as aids in reviewing for an examination. To further help in structuring this review, the enclosed content guideline includes the approximate number of items that can be expected to appear within each major area with the exception of Management and Supervision, Education, Research and Development sections which will not appear at this time.

Please contact the specific staff units within the Board listed below if you have any questions regarding the following topics:

For Questions Concerning:	Contact:
Application Status	Applicant Services (ext. 125, 133, 134)
Reporting of Scores	Applicant Services (ext. 130, 132)
Examination Procedures	Administrative Services (ext. 141, 217)

c/o Board of Registry
P.O. Box 4872
Chicago, Illinois 60680
(312) 738-1336

© 1978 by American Society of Clinical Pathologists
REV. 12/78

EXAMINATION PROCEDURES

EXAMINATION QUESTIONS

All Board of Registry examinations contain written objective questions which may include: multiple choice, matching, multiple response, and in Nuclear Medicine, several true-false questions (see page 3 for examples of these formats). Some of the questions may refer to charts, diagrams, color plates, or other visual material. The Board of Registry incorporates (pretests) several objective questions at random within most its examinations for the purpose of obtaining statistical information. This is done to evaluate the appropriateness of particular questions at the level for which they are intended, thus enabling the examination committees to make a more valid decision for incorporating a question into a future examination. These pretest questions are not used to calculate means, pass/fail points or scores, but only for response feedback.

In addition to the administration of objective questions, the specialist examinations in Chemistry, Hematology, and Microbiology include a series of essay questions; the Blood Banking and Histologic Technic examinations include a practical component and the Cytotechnology examination includes a visual component.

TIME LIMITS

A maximum of four (4) hours is allotted for the examination.

SCORING PROCEDURES

At the present time, the Board of Registry utilizes a norm-referenced procedure for determining minimum passing scores on its examinations. A minimum passing score is calculated by subtracting one standard deviation from the mean. For example, if the mean score on the Medical Technologist examination is 135 points, and the standard deviation equals 20, the passing score would be equal to 115 points. In this example, any examinee attaining a score at or above 115 passes the test and is subsequently certified.

TEST CENTER PROCEDURES

I. When you take your examination, you should bring with you at least two No. 2 pencils with good erasers.

II. Scratch paper will be provided for you. No books, dictionaries, or paper may be taken into the examination room.

III. Though they are not necessary, the Board of Registry does allow the use of slide rules and calculators by applicants. They must be brought in without cases.

IRREGULARITIES

If an examinee is found cheating on a certifying examination, the results will be held until such time as the applicant appeals to the Board of Registry. The Board will review each individual case and determine the appropriate consequences.

REPORTING OF SCORES

The examination results are generally released approximately six (6) weeks following the date of the examination. It is *in*advisable to call the Board of Registry for your results, since test results *cannot* be released over the telephone to anyone. It is suggested that you inform the Board *in writing* if you do not receive your score within two (2) months after taking the test. If you are thus affected, you may contact the Scheduling and Reporting Unit (see first page of memorandum for address).

Individuals who fail an examination will receive, in addition to their scores, information regarding reapplication.

EXAMPLES OF OBJECTIVE QUESTION FORMATS

The following questions illustrate the variety of objective item *formats* that *may* appear on a Board of Registry examination; not all question types will necessarily be administered on any particular examination. The most widely used formats on all examinations are the multiple choice and multiple response.

Please note that the material presented in these questions is not necessarily laboratory relevant nor representative of the content of Board of Registry examinations. These questions are presented *solely* for the purpose of illustrating question *format*.

MULTIPLE CHOICE QUESTIONS

Directions: Each of the questions or incomplete statements below is followed by 5 suggested answers or completions. Select the *one* that is *BEST* in each case and blacken the appropriate space on the answer list.

1. Which of the following U.S. cities is known for its abundance of lakes?

A. Minneapolis
B. Chicago
C. St. Louis
D. Dallas
E. San Francisco

Answer A

MULTIPLE RESPONSE QUESTIONS

Directions: For each of the incomplete statements below, ONE or MORE of the completions given is correct. On the answer sheet blacken the space beside:

A. if only 1, 2, and 3 are correct
B. if only 1 and 3 are correct
C. if only 2 and 4 are correct
D. if only 4 is correct
E. if all are correct

2. Which of the following illnesses is associated with a vitamin deficiency?

1. Scurvy
2. Pneumonia
3. Rickets
4. Hepatitis

Answer B

MATCHING QUESTIONS

Directions: For items 3-6 below, indicate the letter of the most closely related item. The same letter may be used once, more than once, or not at all.

A. 4 players
B. 5 players
C. 6 players
D. 9 players
E. 11 players

3._______Baseball
4._______Basketball
5._______Football
6._______Hockey

Answers 3. D
4. B
5. E
6. C

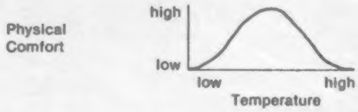
TRUE FALSE QUESTIONS

Directions: Please indicate whether the following statements are true or false. If true, blacken *a* on your answer sheet, if false, blacken *b.*

7. The number of feet in a mile is 1,000. True False
8. The most common blood type is O. True False

Answers 7. B
8. A

EXAMPLE OF QUESTION BASED ON VISUAL MATERIAL



The graph presented above shows the relationship between temperature and an index of physical comfort. On the basis of these data, the BEST conclusion regarding this relationship is that:

A. as the temperature gets warmer, people become more comfortable
B. as the temperature gets cooler, people become more comfortable
C. temperature and physical comfort are not linearly related
D. temperature and physical comfort are both related to a third variable

Answer C

RULES AND REGULATIONS FOR BOARD OF REGISTRY CERTIFICATION

EXAMINATION DATES: Examinations are conducted on the third Friday of February and August each year.

DEADLINE DATES: Your application must be in our office by the first working day of October for the February examination and the first working day of April for the August examination. The Board of Registry strictly adheres to the established deadline dates. Applications received after the deadline will be processed for the following examination.

EXTENDED DEADLINE DATES: Extended deadline dates are indicated in the scoring information for applicants who fail the examination and wish to be re-examined for the next consecutive examination.

APPLICATION FEE: The initial application fee is $50.00. No application will be processed unless accompanied by this fee.

EXAMINATION ELIGIBILITY: Eligibility for examination is established by meeting the current stated minimum requirements for a particular category or level of certification, submission of an application and payment of fee. The data submitted on an application must be verified before an applicant is determined eligible.

OFFICIAL DOCUMENTS:

1. CT and MT students applicants only — A transcript evaluation indicating completion of the required eligibility requirements. Your program should have a copy in your student file. THIS DOCUMENT MUST BE IN OUR OFFICE BEFORE YOU CAN BE SCHEDULED FOR EXAMINATION.
2. MT student applicants only — An official transcript from a regionally accredited college/university bearing the seal of the college/university, the signature of the Registrar and the date your degree was conferred or will be conferred. THIS DOCUMENT MUST BE IN OUR OFFICE BEFORE EXAMINATION SCORES CAN BE RELEASED.
3. All other categories which require a college background — An official transcript from a regionally accredited college/university bearing the seal of the college/university, the signature of the Registrar and the date the degree was conferred. THIS DOCUMENT MUST BE IN OUR OFFICE IN ORDER TO DETERMINE ELIGIBILITY.

CHANGE OF NAME AND/OR ADDRESS: Notify us promptly if you have a change of name and/or address. Changes received within the eight (8) week period prior to the examination date will not be reflected on your examination schedule letter.

CHANGE OF EXAMINATION AREA: Requests for change of examination area must be in our office two weeks before the administration of the examination and must be accompanied by an address change. A fee of $10.00 is charged for this service after schedule letters have been mailed. No transfers will be made within the same testing area.

RELEASE OF SCORES: Examination scores are released to the applicant and the officials of the program in which the examinee was enrolled.

REQUEST FOR HANDSCORE: An applicant is allowed 90 days from the date of receipt of scores to question examination results.

APPEALS MECHANISM: Complaints and appeals must be submitted to the Board of Registry in writing within 90 days of incident or notification. The decision of the Board will be reported to the appropriate individual(s).

RETAKE AND RESCHEDULE FEES: If you cancel, do not appear, fail the examination or are determined ineligible for examination, you must submit a full application fee prior to the *established deadline date along with your request to be rescheduled for examination. (A new application form is not required.)
*Established deadline dates: First working day of April or October.

REFUNDS: If you are determined ineligible for examination, 50% of your application fee will automatically be refunded to you. Upon receipt of a full application fee and notification that all deficiencies have been removed, your application will be processed for the next possible examination. Since applications are retained in our office for a period of five years, you do not need to resubmit an application form.
50% of the application fee will be refunded if you cancel at least five weeks prior to the examination date. After that time, no refunds will be issued.

REEXAMINATION: You are permitted to take the examination a total of six times. If you are unsuccessful after six attempts, you are ineligible for further examination.

APPLICATIONS HELD FOR FIVE YEARS: Applications remain valid for a period of five years from the date of submission. If you wish to be rescheduled for the same category of examination within this time limit, submit a full fee of $50.00 along with the following information: Name (Maiden name when applicable), address, date applied and last examination taken (if applicable). After five (5) years, it will be necessary for you to submit a new application, with the full fee and meet current requirements.

CAHEA ACCREDITED PROGRAMS: The education received from a CAHEA accredited Program is acceptable for a period of five (5) years from the date of completion.

MAIL GENERAL CORRESPONDENCE TO: Board of Registry, P.O. Box 12270, Chicago, Illinois 60612. Address correspondence containing a fee payment to Board of Registry, P.O. Box 97970, Chicago, Illinois 60690. Address correspondence/transcripts to clear HOLD GRADES to: Board of Registry, P.O. Box 5054, Chicago, Illinois 60680.

FACT SHEET

IMPORTANT: Please retain this Fact Sheet containing the list of examination areas and the rules and regulations for your information.

PERMANENT BOARD OF REGISTRY
EXAMINATION AREAS

ALABAMA
Birmingham

ALASKA
Anchorage

ARIZONA
Phoenix

ARKANSAS
Little Rock

CALIFORNIA
Los Angeles
San Francisco

COLORADO
Denver

CONNECTICUT
Hartford

DISTRICT OF COLUMBIA
Washington, D.C.

FLORIDA
Jacksonville
Miami

GEORGIA
Atlanta

HAWAII
Honolulu

IDAHO
Boise

ILLINOIS
Chicago

INDIANA
Indianapolis

IOWA
Davenport

KENTUCKY
Lexington

LOUISIANA
New Orleans

MAINE
Augusta

MASSACHUSETTS
Boston

MICHIGAN
Detroit

MINNESOTA
Minneapolis

MISSISSIPPI
Jackson

MISSOURI
St. Louis
Kansas City

MONTANA
Great Falls

NEBRASKA
Omaha

NEW JERSEY
Edison

NEW MEXICO
Albuquerque

NEW YORK
New York
Albany
Buffalo

NORTH CAROLINA
Durham

NORTH DAKOTA
Bismarck

OHIO
Columbus
Cleveland

OKLAHOMA
Oklahoma City

OREGON
Portland

PANAMA CANAL ZONE

PENNSYLVANIA
Pittsburgh
Philadelphia

PUERTO RICO
San Juan

RHODE ISLAND
Providence

SOUTH CAROLINA
Charleston

SOUTH DAKOTA
Rapid City

TENNESSEE
Knoxville
Memphis

TEXAS
Dallas
Houston
San Antonio
El Paso

UTAH
Salt Lake City

VERMONT
Burlington

VIRGINIA
Richmond

WASHINGTON
Seattle

WEST VIRGINIA
Charleston

WISCONSIN
Milwaukee

WYOMING
Laramie

BOARD OF REGISTRY
EXAMINATION CONTENT GUIDELINES

MEDICAL TECHNOLOGIST EXAMINATION

MICROBIOLOGY KNOWLEDGE AREA

Content Areas I through V have essentially the same general format as appropriate for each Area.

		Number of Questions
I.	Cocci — gram positive and negative	3 – 5
II.	Gram positive, aerobic and anaerobic rods	3 – 5
III.	Enteric and other gram negative rods	4 – 7
IV.	Mycobacteria and miscellaneous organisms (spirochetes, mycoplasma, etc.)	4 – 7
V.	Mycology (yeast, molds and yeast-like molds)	4 – 6

- A. Morphology (cellular)
  - 1. General
    - a. shape
    - b. arrangement
  - 2. Use of stains — staining characteristics
  - 3. Structures
    - a. flagella
    - b. spores
    - c. capsule
    - d. granules
- B. Cultural characteristics
  - 1. Media
    - a. growth requirements
    - b. inhibitory compounds
    - c. role of nutrients in metabolism
  - 2. Redox potential
  - 3. Environmental requirements
  - 4. Colonial appearance
    - a. size, shape, consistency, etc. of colony
    - b. chromogenesis
    - c. effect on blood
    - d. effect on other media
- C. Secondary identification
  - 1. Physiological characteristics
  - 2. Biochemical reactions
  - 3. Antigenic structures and serologic identification
- D. Specimens
  - 1. Sources
  - 2. Processing
    - a. safety procedures
    - b. contaminants
    - c. primary culture
- E. Interpretation and/or trouble shooting

©1978 by the American Society of Clinical Pathologists
Rev. 10/78

- VI. Parasites 4–6
  - A. Morphology
    - 1. Shape
    - 2. Structures
  - B. Specimens
    - 1. Sources
    - 2. Contaminants and artifacts
  - C. Life cycles
- VII. Media preparation and reagents 1–2
  - A. Principles
  - B. Procedures
  - C. Interpretation and/or trouble shooting
- VIII. Disinfection and sterilization 1–2
  - A. Principles
  - B. Procedures
  - C. Interpretation and/or trouble shooting
- IX. Quality Control 1–2
  - A. Procedures
  - B. Principles
  - C. Method comparison and/or development
  - D. Interpretation and/or trouble shooting
- X. Antibiotics including sensitivity testing 2–3
  - A. Procedures
  - B. Principles
  - C. Interpretation and/or trouble shooting
- XI. Genetics — principles 0–1
  - A. Chromosome structure and function
  - B. Mutation, transduction, transformation, conjugation
- XII. Pathogenicity 1–2
  - A. Disease states
  - B. Mechanisms
- XIII. Host Defense 1–3
  - A. Non-specific
  - B. Specific
  - C. Contributing factors
    - 1. Corticosteroids
    - 2. Radiation
    - 3. Antibiotics
    - 4. Disease states
- XIV. Epidemiology 1–2
  - A. Prevention
  - B. Spread
  - C. Nosocomial
- XV. Standard Operating Procedure 0–2
  - A. Recording
  - B. Reporting
  - C. Safety
- XVI. Viruses 2–4
  - A. Nature
  - B. Replication
  - C. Types of infections
  - D. Collection and handling of specimens

HEMATOLOGY KNOWLEDGE AREA

- I. Erythrocytes .... 8 – 15 questions
  - A. Formation
    - 1. Bone marrow
    - 2. Liver
    - 3. Spleen
    - 4. Kidney
    - 5. Endocrine glands
  - B. Function
  - C. Collection
    - 1. Anticoagulants
    - 2. Venous
    - 3. Capillary
    - 4. Bone marrow
    - 5. Special handling
  - D. Enumerative Procedures
    - 1. Manual count
    - 2. Automated count
    - 3. Reticulocyte
    - 4. Indices
    - 5. Correlation of results
    - 6. Quality control
  - E. Basic tests
    - 1. Sedimentation rate
      - a. Wintrobe
      - b. Westergren
    - 2. Hematocrit
    - 3. Fragility
    - 4. Sickle cell
    - 5. Hemoglobin
    - 6. Heinz bodies
    - 7. Sugar water test
    - 8. Ham test
    - 9. Hemoglobin electrophoresis
    - 10. Correlation of results
    - 11. Quality control
  - F. Cytochemistry
    - 1. Wright's stain and variations
    - 2. Iron stain
    - 3. Supravital stain
  - G. Enzymes including Genetics
    - 1. Glycolytic
      - a. Glucose-6-phosphate dehydrogenase (G6PD)
      - b. Pyruvate kinase
      - c. Glutathione (GSH)
    - 2. Non-glycolytic
      - a. Lactic acid dehydrogenase (LDH)
  - H. Morphology
    - 1. Normal
      - a. Peripheral
      - b. Bone marrow
    - 2. Abnormal
      - a. Peripheral
      - b. Bone marrow
    - 3. Inclusion bodies
    - 4. Correlation of results
  - I. Radionuclides
    - 1. Blood volume
    - 2. Schilling test, B_{12}, Folate
    - 3. Red cell survival
  - J. Genetics
    - 1. Hemoglobinopathies
  - K. Pathological states
    - 1. Macrocytic anemias
    - 2. Normocytic anemias
    - 3. Microcytic anemias
    - 4. Hemolytic anemias
    - 5. Polycythemias
    - 6. Pancytopenias
- II. Leukocytes.............................8 – 13 questions
  - A. Formation
    - 1. Spleen
    - 2. Thymus
    - 3. Bone marrow
    - 4. Lymphatic system
  - B. Function
  - C. Collection
    - 1. Anticoagulants
    - 2. Venous
    - 3. Capillary
    - 4. Bone marrow
    - 5. Special handling
    - 6. Lymph node imprint
  - D. Enumerative Procedures
    - 1. Manual count
    - 2. Automated count
    - 3. Differential
    - 4. Eosinophil count
    - 5. Correlation of results
    - 6. Quality control
  - E. Basic Tests
    - 1. Lupus erythematosus (L.E.)
  - F. Cytochemistry
    - 1. Wright's stain and variations
    - 2. Giemsa
    - 3. Peroxidase
    - 4. Periodic Acid-Schiff
    - 5. Sudan black
    - 6. Leukocyte alkaline phosphatase
  - G. Morphology
    - 1. Normal
      - a. Peripheral
      - b. Bone marrow
      - c. Lymph node
      - d. Electron microscopy
    - 2. Abnormal
      - a. Peripheral
      - b. Bone marrow
      - c. Lymph node
      - d. Electron microscopy

  - 3. Inclusion bodies
  - 4. Correlation of results
- H. Genetics
  - 1. Philadelphia chromosome
  - 2. Anomalies
- I. Pathological States
  - 1. Genetic Inherited Conditions
  - 2. Infections
    - a. Bacterial
    - b. Viral
    - c. Parasitic
    - d. Mycologic
  - 3. Metabolic and Toxic Conditions
  - 4. Leukemoid States
  - 5. Neoplastic Diseases
    - a. Leukemias
    - b. Lymphomas
    - c. Multiple myelomas
    - d. Metastases
  - 6. Miscellaneous Conditions of Unknown Etiology
    - a. Lupus Erythematosus
    - b. Pancytopenias

III. Platelets. . . . . 2 – 4 questions

- A. Formation
  - 1. Spleen
  - 2. Liver
  - 3. Bone marrow
- B. Function
- C. Collection
  - 1. Anticoagulants
  - 2. Venous
  - 3. Capillary
  - 4. Bone marrow
  - 5. Special handling
- D. Enumerative Procedures
  - 1. Manual counts
    - a. Counting chamber
    - b. Slide method
    - c. Phase
  - 2. Automated counts
  - 3. Correlation of results
  - 4. Quality control
- E. Enzymes and Factors
  - 1. Adenosine triphosphate (ATP)
  - 2. Platelet factors 1 through 7
- F. Special tests
  - 1. Clot retraction
  - 2. Rumpel-Leede
  - 3. Adhesiveness
  - 4. Platelet factor assay
- G. Morphology
  - 1. Normal
    - a. Peripheral
    - b. Bone marrow
  - 2. Abnormal
    - a. Peripheral
    - b. Bone marrow
  - 3. Correlation of results
- H. Pathological States
  - 1. Thrombocytopenias
  - 2. Thrombasthenias
  - 3. Thrombocythemias

IV. Bone Marrow in Miscellaneous Pathological States . . . . 2 – 4 questions

- A. Morphology
  - 1. Genetic Inherited Conditions
  - 2. Infections
    - a. Bacterial
    - b. Viral
    - c. Parasitic
    - d. Mycologic
  - 3. Metabolic and Toxic Disorders
  - 4. Leukemoid States
  - 5. Neoplastic Diseases
    - a. Leukemias
    - b. Lymphomas
    - c. Multiple myelomas and gammopathies (paraproteinemias)
    - d. Metastases
  - 6. Miscellaneous Conditions of Unknown Etiology
- B. Techniques
  - 1. Specimen Handling
    - a. Tissue Block
    - b. Particle Slide Smears
    - c. Culture
  - 2. Special Stains
  - 3. Chemical Analysis
  - 4. Slide Evaluations

V. Coagulation. . . . . 6 – 8 questions

- A. Theory
  - 1. Intrinsic
  - 2. Extrinsic
- B. Factors
  - 1. Nomenclature
  - 2. Function
- C. Basic Tests
  - 1. Bleeding time
  - 2. Clotting time
  - 3. Plasma recalcification time
  - 4. Prothrombin time tests
  - 5. Activated Partial Thromboplastin Time (APTT)
  - 6. Fibrinogen
  - 7. Thrombin time
- D. Special Tests
  - 1. Substitution tests
  - 2. Factor assays
  - 3. Euglobulin lysis
  - 4. Fibrin Split Products
- E. Quality Control
- F. Correlation of Results
- G. Pathological States
  - 1. Intrinsic
    - a. Hemophiliac States
    - b. Prothrombin Complex Abnormalities
    - c. Fibrinogen, fibrinolysis, Complex Abnormalities

2. Extrinsic
  a. Factor VII
  b. Other Tissue Factors
3. Combined Deficiency States
  a. von Willebrand's Disease
  b. Disseminated Intravascular Coagulation (DIC)

VI. Instrumentation .... 2–4 questions
  A. Principles
  B. Components
  C. Standardization
  D. Preventive Maintenance
  E. Trouble Shooting

VII. Laboratory Safety .... 1–2 questions

VIII. Miscellaneous .... 1–2 questions
  A. Spinal Fluid Cell Count and Differential

IMMUNOLOGY (IMMUNOHEMATOLOGY) KNOWLEDGE AREA

I. General — The Immune Process .... 11 – 15 questions
  A. Natural immunity and immunologic substances
    1. Lysozymes
    2. Basic polypeptides
    3. Properdin
  B. Antigens
  C. Humoral and Cellular immunity
    1. Specificity
    2. Mechanism of antibody formation
      a. Lymphocyte function
      b. Thymus gland function
    3. Immunoglobulins
      a. Classification
      b. General properties, structure
      c. Identification techniques
        1) Immunoelectrophoresis
        2) Immunodiffusion
        3) Immunofluorescence
        4) Radioimmunoassay (RIA)
    4. Primary and secondary response
    5. Passive immunity
  D. Antigen — antibody reactions
    1. Indicators of antigen-antibody reactions
      a. Agglutination
      b. Lysis, including hemolysis
      c. Precipitation
      d. Complement fixation
      e. Fluorescent Antibody Technique
    2. Neutralization of antigens by antibodies
    3. Opsonization
  E. Complement
  F. Autoimmunity
    1. Natural immunologic tolerance of one's own protein (self concept)
    2. Autoimmune responses
      a. Autoimmune hemolytic anemias
        1) Acquired hemolytic anemias caused by incomplete warm antibodies
        2) Acquired hemolytic anemias caused by cold autoantibodies
        3) Paroxysmal cold hemoglobinuria
        4) Atypical autoimmune hemolytic anemias
      b. Autoimmune collagen disease
    3. Induced tolerance, graft rejection and tumor immunology
  G. Allergy — delayed hypersensitivity

II. Serologic Tests for Syphilis .... 2 – 4 questions
  A. General principles
  B. False positives
  C. Flocculation test
  D. Complement fixation tests
  E. Fluorescent Treponema Antibody (FTA)
  F. Quality control
    1. Reporting results
    2. Record Maintenance
    3. Reagent control
    4. Equipment maintenance

III. Serodiagnostic Tests — Non-syphilitic .... 9–11 questions
  A. Widal reaction
  B. Opsonocytophagic index
  C. Febrile agglutinins
    1. Brucellosis
    2. Weil-Felix (Proteus OX19, OX2, OXK)
    3. Paratyphoid A
    4. Typhoid H
    5. Salmonella A-E
  D. Cold agglutinins including streptococcus MG agglutination
  E. Streptococcus grouping
  F. Serologic tests for rheumatoid arthritis
    1. Latex fixation tests (including screening)
    2. Bentonite flocculation
  G. Antistreptolysin O titers
  H. Antihyaluronidase titers
  I. C-reactive protein
  J. Autoimmune substances
  K. Infectious Mononucleosis Testing
    1. Heterophile
      a. Presumptive
      b. Differential
    2. Monospot
  L. Pregnancy Tests
  M. Other
    1. Tularemia
    2. Leptospirosis
    3. Toxoplasmosis
    4. Echinococcus
    5. Trichinosis
  N. Quality control
    1. Reporting results
    2. Record maintenance
    3. Reagent control
    4. Equipment maintenance

IV. Immunohematology — Blood Banking .... 12–14 questions
  A. Donor selection; Processing of blood; Blood collection
    1. Donor identification
    2. General qualifications
    3. Medical indications for donor rejection
      a. Protection of donor
      b. Protection of recipient
    4. Physical exam including lab data
    5. Therapeutic bleeding
    6. Autologous transfusions
    7. Plasmapheresis
    8. Routine donations
      a. Blood collection
      b. Blood processing and labeling
      c. Blood preservation and storage
    9. Donor reactions
  B. Blood components — preparation, and indication for use
    1. Whole blood
    2. Packed cells
    3. Fresh frozen plasma
    4. Cryoprecipitate
    5. Platelet rich plasma
    6. Leukocyte poor blood

C. Preparation and administration of blood
  1. Patient identification
  2. Blood grouping and typing
  3. Crossmatch (major and minor)
    a. Saline, room temperature phase
    b. High protein, 37 °C phase
    c. Enzyme
    d. Antihuman Globulin (Coombs)
  4. Antibody detection and identification
    a. Screening cells, cell panels
    b. Titration
    c. Absorption, elution
    d. Factors affecting blood group antigen-antibody reactions
  5. Problem solving in crossmatches
  6. Emergency use of blood
  7. Transfusion reaction

D. Other general blood banking considerations
  1. Genetics of blood group inheritance
  2. ABO groups
  3. Rh-Hr system
  4. Other blood group systems
  5. Hemolytic disease of newborn
    a. Cause — Rh, ABO, other
    b. Detection
      1) Elution
      2) Antibody identification
    c. Exchange transfusion
    d. Therapeutic uses of Rh immune globulin

E. Quality control
  1. Reporting results (grouping, typing, crossmatching)
  2. Release and transfer of blood and components
  3. Record maintenance
    a. Pre-transfusion
    b. Post-transfusion
  4. Reagent control
  5. Equipment maintenance

MISCELLANEOUS TOPICS

This outline covers the following knowledge areas: Urine & Other Body Fluids, Cytogenetics, Management & Supervision, Education, Research & Development and Fundamental Analytical Principles and Laboratory Calculations.

URINE & OTHER BODY FLUIDS (16 – 20 questions)

I. Renal Physiology .... 2 – 4 questions
    A. Glomerular Function
    B. Tubular Function
        1. Secretion
        2. Reabsorption
        3. Threshold
    C. Endocrine
    D. Volume
    E. Urine Composition
    F. Rate of Formation
    G. Correlation of Results with Pathophysiology
II. Routine Urinalysis .... 9 – 13 questions
    A. Physical Tests
    (includes principles & sources of error)
        1. Appearance
        2. pH, titratable acidity
        3. Concentration
            a. specific gravity
            b. refractive index
            c. osmolality
    B. Chemical Tests
    (includes principles & sources of error)
        1. Carbohydrates
            a. glucose
            b. reducing sugars
            c. other reducing substances
        2. Proteins
        3. Ketones
        4. Pigments
            a. bilirubin
            b. urobilinogen
            c. hemoglobin, myoglobin
            d. homogentisic acid, melanin
            e. porphyrins
        5. Miscellaneous
            a. phenylpyruvic acid
            b. urinary calculi
    C. Sediment Examination
        1. Cells
        2. Casts
        3. Crystals
        4. Lipids
        5. Parasites, fungi, bacteria
    D. Specimen Collection & Handling
    E. Quality Control
    F. Correlation of Test Results
III. Other Body Fluids .... 3 – 7 questions
    A. Cerebrospinal Fluid
        1. Physiology
        2. Specimen collection & handling
        3. Correlation of test results
    B. Feces
        1. Physiology
        2. Specimen collection & handling
        3. Physical tests
        4. Chemical tests, qualitative
            a. occult blood
            b. fat
            c. pepsin
        5. Correlation of test results
    C. Gastric Fluid
        1. Physiology
        2. Specimen collection & handling
        3. Gastric analysis
        4. Correlation of test results
    D. Seminal Fluid
        1. Physical tests
        2. Specimen collection & handling
        3. Fertility & sterility studies
        (including routine semen analysis)
        4. Medical-legal studies
        (typing, chemistry, etc.)
        5. Correlation of test results
    E. Amniotic Fluid
        1. Physical tests
        2. Specimen collection & handling
        3. Correlation of test results
    F. Ascitic Fluid & Pleural Fluid
        1. Physiology
        2. Specimen collection & handling
        3. Correlation of test results
    G. Duodenal Fluid
        1. Physiology
        2. Specimen collection & handling
        3. Correlation of test results
    H. Salivary Fluid
        1. Physiology
        2. Specimen collection & handling
        3. Correlation of test results
    I. Synovial Fluid
        1. Physiology
        2. Specimen collection and handling
        3. Correlation of test results

CYTOGENETICS (1 – 3 questions)

I. Cytogenetics
    A. Genetic principles
    B. Specimen collection & processing
        1. Examination of smears
        2. Cell culture techniques
        & analysis
    C. Correlation of results

MANAGEMENT & SUPERVISION (3 – 5 questions)

I. Budget
  A. Operating Expenses
  B. Cost Accounting
  C. Salary Expenses
  D. Justification of Change in Budget
II. Personnel
  A. Work load analysis – scheduling
    1. Number of tests
    2. Number of hours
  B. Hiring Practices
  C. Advancement Practices
  D. Termination Practices
  E. Evaluation of Personnel
III. Laboratory Space
  A. Planning New Facilities
  B. Utilization of Existing Facilities
IV. Laboratory Supplies
  A. Ordering
  B. Storing
  C. Inventory
V. Laboratory Equipment – Fixed or Heavy
  A. Procurement – Purchase, Lease
  B. Maintenance, Including Record Keeping & Quality Control
  C. Trouble Shooting, Including Record Keeping & Quality Control
VI. General Principles of Records, Federal & State Regulations, and Laboratory Safety
  A. Records
  B. Federal & State Regulations
  C. Laboratory Safety (includes radiation safety)
VII. Ethics & Medical/Legal Matters
  A. Ethics
  B. Medical/Legal Matters

EDUCATION (3 – 4 questions)

I. Job Entry Indoctrination
II. Continuing Education for Laboratory Employees
III. Interdepartmental Rotation
IV. Inservice Education
  A. Interdepartmental
  B. Intradepartmental
V. Effective Mechanisms of Teaching
  A. Objectives
  B. Curriculum Development
  C. Evaluation

RESEARCH & DEVELOPMENT (1 – 3 questions)

I. Evaluation of Test Performance
  A. Quality
  B. Quantity
  C. Cost
II. Initiation of New Procedures
  A. Criteria for Initiation
  B. Evaluation Criteria
    1. Accuracy
    2. Precision
    3. Cost
  C. Information Dissemination
    1. To laboratory personnel
    2. To physicians
    3. Change of normal values
    4. To other institutional personnel departments

FUNDAMENTAL ANALYTICAL PRINCIPLES AND LABORATORY CALCULATIONS (10 – 14 questions)

I. Solutions and Standards .... 1 – 3 questions
  A. Molal, Molar, Normal, and Percent Solutions
  B. Specific Gravity
  C. Hydrates and Water of Hydration
  D. Dilution of Solution and Serial Dilution
  E. Conversion Problems
  F. Balances and Weighing
  G. Analytical Standards
  H. Glassware precision
II. Preparation of Graphs..................... 0 – 1 questions
  A. Use of Graph Paper (Cartesian, Log, Semilog)
  B. Coordinates
  C. Standard Curves
  D. Units
III. Spectrophotometric Calculations .... 1 – 3 questions
  A. Applications of Beer's Law
  B. Molar Extinction Coefficient
IV. General Principles of Quality Control .... 3 – 5 questions
  A. Basic Statistics
    1. Mean, median, mode
    2. Standard deviation, variance
    3. Coefficient of variation
  B. Quality Control
    1. Accuracy, precision, reliability, significant figures
    2. Confidence limits, percentiles
    3. Distribution characteristics
    4. Development of normal values
    5. Control charts
V. Hydrogen Ion Concentration & pH .... 2 – 4 questions
  A. Standard Acids and Bases
  B. pH Reference Solutions
  C. Calculation of Hydrogen pH
  D. Calculation of Hydrogen Ion Concentration
  E. Buffers, Henderson-Hasselbalch Equation
VI. Principles of Microscopy .... 0 – 1 questions
  A. Light
  B. Dark Fields
  C. Fluorescence
  D. Phase Contrast
  E. Interference
VII. Principles of Nuclear Medicine .... 0 – 1 questions
  A. Radionuclide Physics
  B. Regulations & Safety

CHEMISTRY KNOWLEDGE AREA

I. Instrumentation — The following classes of instruments are to be covered.................................. 5 – 7 questions
   1. Coulometric
   2. Spectrophotometric & Photometric
      a. emission flame
      b. atomic absorption
      c. colorimeter
   3. Fluorescence & Phosphorescence
   4. Automated Analyzers
      a. continuous flow
      b. discrete sampling
   5. Electrophoretic & Chromatographic
   6. Potentiometric (including blood gas analyzers)
   7. Recording devices
   8. Centrifuges
   9. Osmometers
   10. Balances
   11. Radionuclide instruments
  A. Theory of Operation
     1. Physical principles involved in operation
     2. Chemical principles involved in operation
  B. Essential Components of Instruments —
     Example: Colorimeter
     1. Power supply
     2. Exciter lamp
     3. Monochromator
     4. Sample holder
     5. Photo detecter
     6. Readout device
  C. Mechanism of Operation
  D. Application/Use of Instruments
     1. Operation
     2. Maintenance
        a. calibration records
        b. component changes & documentation
     3. Choice of Instrument Type
  E. Special Precautions, Specimen Collection & Trouble Shooting

II. Theory & Application of Physiologic Biochemistry.................................. 33 – 35 questions
  The following subtopics and descriptive components will be applied to topics A-J.
   1. Normal & abnormal physiology
   2. Test procedures
      a. principle
      b. procedure
      c. special precautions/specimen collection & trouble shooting
   3. Interpretation
      a. related to physiology
      b. normal vs. abnormal
      c. validity & accuracy of procedure
      d. interrelationship with other laboratory tests
   4. Quality control
      a. statistics & data recording
      b. interpretation & use of data
   5. New test development
      a. evaluation of new procedures
      b. mechanics of setting up new procedures

A. Carbohydrates: .... 3 – 5 questions
To include all carbohydrates of medical importance whether monosaccharides, disaccharides or polysaccharides. Identification tolerances, etc.

B. Electrolytes: .... 3 – 5 questions
To include the major electrolytes and trace elements. Substances such as amino acids and proteins are covered seperately in topic E.

C. Acid Base Balance/Blood Gases: .... 3 – 5 questions
Distribution of ions in intracellular and extracellular compartments, Gibbs-Donnan Equilibrium and osmolality; Henderson-Hasselbach equation for bicarbonate-carbonic acid systems; other body buffer systems; transport of oxygen and carbon dioxide.

D. Enzymes: .... 3 – 5 questions
To include all enzymes of clinical significance. Enzyme kinetics, factors governing the rate of enzyme reactions, enzyme activators and inhibitors, enzyme denaturation and inhibition, role of coenzymes and isoenzymes.

E. Proteins & Amino-Acids: .... 3 – 5 questions
To include all simple and complex proteins except lipoproteins and enzymes.

F. Nonprotein Nitrogenous Substances: .... 3 – 5 questions
To include all major components such as creatinine, urea nitrogen and uric acid.

G. Heme Derivatives: .... 3 – 5 questions
To include heme precursors such as pyrroles and porphyrins, heme per se, and the metabolic breakdown products related to heme such as bilirubin and urobilinogen.

H. Lipids: .... 3 – 5 questions
To include the following: lipoproteins, non-esterified fatty acids, triglycerides, glycerophosphatides, sphingolipids, sterols and fatty acids.

I. Endocrinology: .... 3 – 5 questions
To include the nature and action of hormones, control of hormone secretion. Specific areas to be covered include adrenocortical hormones, steroids, androgens, estrogens, prostaglandins, pituitary hormones, insulin, catecholamines, vanilmandelic acid, serotonin and 5-hydroxyindole acetic acid.

J. Toxicology: .... 3 – 5 questions
To include gases such as carbon monoxide; volatile substances such as ethanol, methanol, cyanide, kerosene; corrosive metals such as arsenicals, lead; non-metals such as bromides, fluorides; organic substances such as amphetamines, barbiturates, salicylates.

FORM No. F.C.4

EXAMINATION REPORT

NAME and ADDRESS

SAMPLE

EXAMINATION DATE
AUGUST 17, 1979
CONTROL NUMBER
999006969
TEST TYPE
MT

CATEGORY	MAX. SCORE	YOUR SCORE
BACTERIOLOGY	29	21
BLOOD BANKING	17	13
CALCULATIONS	19	14
CHEMISTRY	40	24
HEMATOLOGY & COAG	36	30
PARASIT & MYCO	10	08
SEROLOGY	23	17
URINALYSIS	21	16
MISCELLANEOUS	05	04
TOTAL	200	147

PASSING SCORE 128

PASS PASS

CERTIFICATION NUMBER MT 222222

AMERICAN SOCIETY OF CLINICAL PATHOLOGISTS
P. O. Box 4872 • Chicago, Illinois 60680 • Phone (312) 738-1336

DEAR CERTIFICANT:

IT IS A PLEASURE TO INFORM YOU THAT YOU PASSED THE BOARD OF REGISTRY EXAMINATION ADMINISTERED ON AUGUST 17, 1979 . WE WELCOME YOU TO THE RANKS OF OUR CERTIFICANTS. YOU ARE NOW ENTITLED TO USE MT(ASCP) AFTER YOUR NAME, INDICATING THAT YOU ARE CERTIFIED BY THE BOARD OF REGISTRY OF THE AMERICAN SOCIETY OF CLINICAL PATHOLOGISTS. YOUR CERTIFICATE AND IDENTIFICATION CARD ARE ATTACHED.

IT IS ESSENTIAL THAT YOU KEEP US INFORMED OF YOUR CORRECT ADDRESS. THE ADDRESS SHOWN ON THE IDENTIFICATION CARD IS RECORDED ON OUR DATA PROCESSING RECORDS. IF THIS INFORMATION IS INCORRECT, OR IF YOU CHANGE YOUR NAME OR ADDRESS PLEASE NOTIFY US PROMPTLY, STATING YOUR FULL NAME, CERTIFICATION NUMBER, OLD AND NEW ADDRESSES, INCLUDING ZIP CODES. YOUR CERTIFICATION NUMBER IS INCLUDED IN THE ADDRESS BLOCK ON YOUR IDENTIFICATION CARD AND SHOULD ALWAYS BE INCLUDED IN COMMUNICATIONS WITH THE BOARD OF REGISTRY.

CONGRATULATIONS AND BEST WISHES FOR YOUR CONTINUED SUCCESS.

SINCERELY YOURS,

Barbara M. Castleberry

Barbara M. Castleberry, M.S., MT(ASCP)
CHAIRMAN, BOARD OF REGISTRY

SAMPLE

JACK RAINBOLT
MT(ASCP) 222222
200 E. ONTARIO STREET
CHICAGO, IL 60611

YOUR IDENTIFICATION CARD. CUT ALONG DOTTED LINES.

PREPARED STATEMENT OF DR. OWEN B. KIERNAN, EXECUTIVE DIRECTOR, NATIONAL ASSOCIATION OF SECONDARY SCHOOL PRINCIPALS

Mr. Chairman and distinguished members of the Committee, I thank you for the opportunity to share with you the National Association of Secondary School Principals' concerns about a bill currently before your Committee, the Educational Testing Act of 1979 (H.R.4949). First, you should know that the NASSP represents more than 35,000 secondary school administrators nation-wide, who in turn are responsible for the education of 20 million American youth.

We have reviewed this proposal, and wish to explain our firm opposition to it on the grounds that (1) it calls for the federal government to interfere with, and intrude into, an important area of education in a manner which is totally inappropriate; (2) it represents yet another example of federal paperwork to which Americans are taking increasing exception; and (3) it enlists Uncle Sam in federalizing admission policies in our higher education institutions. Let me be more specific in our opposition to what we believe to be a bad bill.

First, it has long been our belief that education is a *local function, a state responsibility, and a federal concern.* Only two weeks ago, Congress gave its final approval to legislation creating a separate cabinet-level Department of Education. We applaud this action. In that long and heated debate the Congress, in its wisdom, included language to forbid federal control of education, and more positively to ensure state and local control of America's schools. H.R.4949 flies in the face of this clear direction, containing the roots of federal intrusion into the highly technical standardized test industry. And which federal agency will be charged with the responsibility to carry out such a law--the Department of Education!

Regarding this intrusion by the federal government, let me cite several glaring examples in the text of H.R.4949. In Section 2(b)4, you will note one of the purposes of the Act is "to encourage use of multiple criteria in the grant or denial of any significant educational benefit." What form will this "encouragement" take? Does this imply that our higher education institutions will be "encouraged" to modify or revise their admissions policies? And how would they be so "encouraged?" This seems to us to be a crude invitation to bureaucratic interference with the legitimate authority of educational administrators, and we vehemently oppose such a purpose.

Furthermore, let me call your attention to Section 7(2)E,F and G where you will note that the Commissioner (Secretary of Education) is empowered to monitor the fiscal activity of nonprofit testing organizations. This is a clearly inappropriate role for the Department of Education to assume. It assigns regulatory functions to the Department in the area of private enterprise which have never been contemplated for this agency. How would the Department use this information regarding revenues and expenses of test-making agencies?

Would it coordinate its information with the Internal Revenue Service? This portion of the bill more than any other smacks of a witch hunt that is absolutely unsupportable, and raises questions as to the real purpose underlying the entire bill. These purposes are presumably stated in Section 2, and have been presented elsewhere as being primarily to protect persons taking standardized tests against error and improper application. Some of the controls and limitations on the use of standard testing sought to be imposed by this bill suggest more far-reaching objectives, however, including the possible elimination of standardized testing entirely on the basis of some philosophical objections not stated in the legislation. Indeed, Subsection (a) 4 of Section 2, setting forth the bill's findings and purpose does say that "there is increasing concern among citizens, educators, and public officials regarding the appropriate use of standardized tests in the admissions decisions of postsecondary education institutions." We would respectfully point out that there is at least as much concern among many citizens, educators, and public officials that reasonable standards be maintained by those institutions to assure the public that the diplomas and degrees they grant are worth more than the paper on which they are printed.

Mr. Chairman, I think it is incumbent upon the members of this Committee to recognize the serious crisis of confidence into which we have fallen in recent years regarding the quality of education at all levels. Without attempting to debate this wider issue here today, it would hardly seem prudent at this time to weaken public confidence further by seeking to destroy mechanisms for assuring some measure of evaluating the quality of student abilities and student achievement, as well as diagnosing weaknesses and providing remedial instruction. To the extent that this is the objective of this and similar legislation, it must be seen as particularly pernicious.

Finally, we oppose H.R.4949 as a further contribution to the paperwork blizzard. We agree that a greater understanding of standardized tests is a right of all test subjects, since their academic futures will be at least partially affected by their performance on such tests. At the same time, this bill exemplifies an age-old federal problem of information overkill. Section 3, for example, requires provision in "understandable language" of the intents, purposes, and skills measured by tests, and an explanation of interpretation of scores. In addition, however, it requires that the subjects of the tests be informed of the "standard error of measurement" and "correlation between scores and success in careers." We contend that these kinds of technical test analyses, although very useful in upgrading tests, will be of little use to test subjects. Instead, such requirements can only be viewed as a continuation of the ill-advised tendency of Congress and federal agencies to collect data for the sake of collecting data. We believe any such procedure to be inappropriate.

The release of test questions and answers as required in Section 5(1) and (2) would not only add to the cost of developing tests but will be of little use to test subjects. It goes without saying that the major effect of releasing questions and answers to tests will not be the improvement of a subject's subsequent score, but merely an increase in the number of sample questions presently supplied. The beneficiaries of released questions and answers may very well be those who run costly test preparation courses rather than the subjects themselves.

In 1978 a Resolution was unanimously adopted by our Association which states, "The NASSP supports standardized testing in secondary schools as an appropriate measure of student performance when such tests *at the state level* are properly devised, administered and interpreted. NASSP encourages continuing efforts to improve standardized testing as opposed to testing abandonment."

This continues to be our position today.

We believe that the approach of H.R. 4949 to the legitimate goal of providing students protection against error and abuse of standardized testing is ill-advised. We urge the Committee to defeat this misguided attempt to bring the standardized test industry into submission.

Thank you for the opportunity to present our views on this legislation.

PREPARED STATEMENT OF THE AMERICAN EDUCATIONAL RESEARCH ASSOCIATION

--Preamble--

The American Educational Research Association (AERA) is a professional association of nearly 14,000 members engaged in the breadth and depth of educational inquiry. Our members are concerned with both the quality (reliability, validity, and accuracy) and the effective and fair administration and use of educational tests. AERA members are actively engaged in the improvement of educational tests and in studies on the educational, psychological, and societal impact of testing.

We submit this statement in the spirit of assisting the Congress in its deliberations and decisions concerning testing legislation which has far reaching implications for the future practice of educational measurement. We believe that the Congress should legislate in this area with great care and, on balance, the services provided to individuals and society by admissions tests outweigh disadvantages inherent in any effort to assess human potential. It is not, however, our intent to take a polemic position on either of these bills or specific aspects of the legislation before the Committee. In fact, there is no clear consensus among researchers and scholars regarding the need for this legislation. There is, however, agreement that the consequences of the proposed legislation are not well known at this time.

We share with Congressmen Weiss, Gibbons, and their co-sponsors the interest in increasing the understanding of the purposes, development procedures, contributions, and limitations of admissions tests. We also share their interest in constructing and applying such tests in a manner which does not arbitrarily discriminate against test subjects because of their race, sex, ethnic, income, or family background. We recognize that individuals from varying backgrounds score differently on admissions tests; however, these differences have also provided needed information to the public and Congress which has been useful in creating an increased awareness of discrimination and prejudice in the society. This increased awareness has been a significant component behind the enactment of programs and has often used the results of educational measurement and ability assessment to argue for enactment of educational legislation.

Educational measurement, including admissions tests, is far from perfect and is limited in the ability to predict future education or occupational performance. As educational researchers and concerned citizens, we hope that any congressional action in the testing arena will improve the development and administration of admissions tests and enhance the equitable and fair treatment of test subjects. We believe that the proposed legislation has already stimulated health discussion and debate. We are, however, not sure that the specific provisions of these bills will accomplish the laudable legislative purposes which they set forth. Legislative action should promote increased understanding of educational measurement by stimulating more and better research and analysis of tests; however, no such action should undermine the validity and reliability of the tests.

The most serious concern with the proposed legislation is the required test item disclosure. Any such measure should improve test development, administration, use, and understanding without providing an unfair advantage to those who can most afford to use the test information.

ACTING FROM A BASE OF KNOWLEDGE

The Committee has heard from numerous witnesses explicating the advantages, potential harm, and unknown consequences of the proposed legislation. We believe that the most salient feature of such testimony has been in the area of "unknown consequences." We, therefore, believe that the Committee's deliberations and decisions would be greatly enhanced with increased knowledge and understanding from the following phenomena:

1. Ability Testing Report. The Committee on Ability Testing of the National Academy of Sciences (see ATTACHMENT -A- for members & staff) has undertaken a "major examination of the role of standardized testing in American society" (including considerations of aptitude, intelligence, achievement, or ability tests)..."It is expected that the (NAS) Committee's work will result in a report that illuminates the problems posed by the use of standardized tests and recommends appropriate policy alternatives." (NAS, March, 1979) The report is expected to be distributed by August 1980.

2. Standards for Testing. AERA, the American Psychological Association, and the National Council on Measurement in Education have agreed to undertake a revision of their Standards for Educational and Psychological Tests. These Standards are widely used. They are recognized in the federal Uniform Guidelines in Employee Selection Procedures and have been used in at least two instances by the U.S. Supreme Court as the standard guide to sound professional practice in testing and by major test publishers throughout the country. The current and revised Standards will provide the Committee with useful background and information on salient issues and topics being considered by researchers, test producers, test administration agencies, and test score users.

3. New York. The recently enacted New York State admissions testing legislation provides a mandate similar to the test disclosure provisions in H.R. 4949. Observation and analysis of the impact of the New York law will provide the Committee with useful information and understanding of the potential consequences of test item disclosure and other specific provisions of the proposed bills.

4. Study Commission or Mandated Study. Alternative #1. Several witnesses have suggested that a congressionally established study commission would signficantly assist the Committee. Alternative #2. The Congress could mandate an admissions testing study to be conducted by National Institute of Education (NIE). The commission/study on educational and occupational admissions testing could use available information from the above phenomena. Additional research and data on testing, which could be provided by the testing agencies, individual research efforts, and federal agencies could also be examined. In addition to providing the Committee with useful, policy relevant information, the commission/study could point out areas where existing knowledge is unavailable or uncertain.

The commission or advisory group for the mandated study could include representatives of various interested parties, such as: test takers, parents, teachers, public interest groups, test producers, test administrators, test users, test development and education measurement experts, researchers (testing and social & individual impact of tests), legal scholars, minorities, women, school and higher education administrators (including professional schools). Ex officio members of Congress and federal agencies concerned with education and occupational testing would provide an ongoing interchange of information while the study and analysis is being conducted.

ISSUES AND A CHARGE TO THE "COMMISSION" OR "STUDY"

The proposed legislation and testimony presented to the Committee provide a useful set of issues and questions for a congressionally charged study or commission. Through a process of interim and final reports, the Committee could benefit from a thorough examination of the following issues:

> A. Is federal legislation needed to assure that test takers have adequate information both prior to and following their participation in admissions tests?

Proponents argue that test agencies either do not provide adequate information or that they only do so when under pressure from outside consumer and public interest goups. They also argue that such information could reduce myths and overreliance on test scores in the admissions process.

Opponents argue that nearly all of the information required by the proposed legislation is already provided and that students, schools, and institutions of higher education already have information related to standard errors of measurement, scoring procedures, limitations, and purposes of the tests.

CHARGE: The commission/study could provide an external validation of the claims made by test agencies and outline for the Committee any discrepencies between information required by the proposed legislation and that already provided to test takers and user institutions and organizations. There could also be an analysis of the degree to which such information is used by test takers and test score users to improve the individual and public understanding of the value and limitations of admissions tests.

> B. Should legislation mandate the disclosure of specific test items used for scoring purposes (including correct answers) to test takers and the federal government, user institutions, and public?

Proponents argue that it is a form of just due process for test takers to see the instruments upon which their academic future will be judged. Such disclosure would also allow a more thorough examination of test validity, bias, and characteristics of admissions tests. They also argue that disclosure would improve test performance given a better understanding of testing domains and areas where participants do poorly. Disclosure would also, according to the proponents, lead to improved test development, expose scoring errors, and improve research on testing.

Opponents argue that test agencies already provide sample tests and that item disclosure will increase costs (to participants), reduce test administrations (particularly for special groups like the handicapped), reduce test reliability and fairness, cause "equating" problems, reduce the value of copyrighted material (without compensation), cause some test producers to cease such activity (causing increased monopoly of test development), increase teaching to the test (thereby increasing the influence of tests on education curricula), and have adverse effects on those test takers least able to afford test coaching or use test items to improve performance. The test agencies argue that they already have elaborate safeguards to prevent scoring errors.

CHARGE: The commission/study could examine several of these issues, including the potential impact on different populations. For example, it would be useful to know if high schools in large cities and remote rural areas would have the capacity to use test items in a manner similar to those in affluent communities with large counseling and guidance staffs. Cooperation with testing agencies, who could provide secure tests and analysis of development procedures and costs, would allow a better understanding of the potential benefits and costs (including validity, reliability, and equating problems) under conditions of disclosure. The New York case will provide valuable insights into many of the unknown consequences.

C. Will the proposed legislation provide an opportunity to clarify the issues surrounding cultural or age, sex, race, or regional bias in admissions tests?

Proponents argue that the tests are inherently biased and have provided the Committee with research studies demonstrating that different vulnerable groups score lower than the national average on some of the most commonly used tests. They claim that these tests play a significant role in limiting the educational and occupational life chances of certain populations. They also claim that the test agencies refuse to reveal information on differential scoring by racial and economic groups.

Opponents do not dispute the differences in test scores among certain groups; however, they argue that such differences are a reflection of the broader society and have provided policy makers with needed information to develop and enact policies which are designed to reduce scoring gaps. Test agencies argue that they engage in research and development efforts to eliminate "arbitrary" bias in the tests. They also state that some data on income and ethnic background is not collected or, if revealed, would be a breach of privacy or agency policy.

CHARGE: The commission/study could examine the available research on test bias. Under agreement to retain test security, they could examine or commission qualified researchers (under the same security agreement) to examine test items used in different tests. Procedures used by test agencies to eliminate or reduce arbitrary bias could be examined and assessed. The commission/study could establish guidelines for determining who should have access to test items for legitimate research purposes.

D. Will the proposed legislation place the use of test in proper perspective with regard to academic and occupational admissions?

Proponents argue that there is an overreliance on test scores in admissions procedures, such procedures are too rigid, the tests do not measure important criteria for academic and occupational performance and success (e.g., motivation, interest, perserverance, etc.), and that the economic benefits or producing cheap paper and pencil tests places an undue emphasis on rote learning to the neglect of conceptual skills.

Opponents argue that the test agencies advise academic institutions that the tests should not be the sole criteria for entry and postsecondary education associations have stated that their members use a variety of criteria beyond test scores to determine academic entry. They also argue that the test producers have eliminated a great deal of bias found in early versions of the tests and that these tests have helped reduce arbitrary selection procedures in academic admissions. Institutions of higher education claim that the proposed legislation could reduce the use of standardized tests and increase the use of "institution made tests" which would likely be of poorer quality. Some fear that the disclosure provisions of the legislation may increase the relative importance of testing beyond stated limitations.

CHARGE: The commission/study could examine the available research on admissions practices and, in cooperation with funding agencies, examine a sample of such institutional practices. They could also examine the scope of the tests and ability of test producers to assess some of the criteria mentioned above.

E. Would the proposed legislation increase the understanding of the impact of test coaching schools?

Proponents argue that coaching does have a positive effect on test performance for some test takers and that test coaching is a growing business. They argue that coaching schools gain access to tests, and for those able to afford them, create an unfair advantage for higher income groups.

Opponents argue that the evidence on test coaching is far from conclusive and that the proposed legislation will create an undue bonanza for coaching schools.

CHARGE: The commission/study could examine the test coaching arena, including an examination of the numbers, types, and effects of coaching schools. Such an examination should also include an assessment of the degree to which test security can be or is being violated by such schools and the ability to coach for the tests.

F. Is federal legislation needed to reveal the financial arrangements between test producing and test administration agencies and the costs and income of test agencies?

Proponents argue that testing is a large industry dominated by a few organizations and corporations. As such, they argue that test agencies should be treated like "public utilities" which are subject to certain financial disclosure requirements. They also argue that such disclosure will reveal the margin of income over cost among the non-profit test agencies. It is argued that such information would be useful in determining the need for increased costs to test takers and users should the reporting and item disclosure requirement of the legislation be enacted.

Opponents argue that test agencies provide a valuable public service. Nonprofit test agencies state that they use income for a variety of purposes designed to improve testing, including research and development projects on tests. Both commercial and nonprofit test agencies say that the proposed financial disclosure requirements are not based on substantiated claims of misuse of the funds they receive. They also say that this disclosure would place them at a disadvantage in competition for contracts and that certain rights of privacy will be violated by some of the disclosure provisions.

CHARGE: The commission/ study could examine the legal and constitutional issues surrounding the various financial disclosure provisions of the proposed legislation. Since such requirements are not included in the New York law, such an examination would be valuable for the Committee to determine the feasibility of enforcement and the legal problems associated with financial disclosure.

G. Is federal legislation needed to stimulate improved adherence to professional standards of test development and use?

Proponents argue that the cultural bias, secrecy, and limitations of the admissions tests indicate that professional standards are not followed, not enforced, or that it is impossible to determine if they are followed.

Opponents argue that test agencies follow the requirements of the AERA, APA, NCME Standards and that no charges of violation have been substantiated. Test agencies state that they also adhere to their own guidelines and procedures designed to assure fairness and proper development.

CHARGE: The commission/study could conduct an examination of the degree to which such Standards and other codes of testing conduct are followed.

H. Is legislation necessary to increase the federal study of admissions testing and to provide increased information to the public and researchers who otherwise do not have access to such data and information?

Proponents argue that there is an inadequate base of knowledge and information to conduct needed research on testing, test development and the effects of tests on different populations. They argue that attempts to gain study data on cultural bias and test validity and reliability is either difficult to obtain or not available (either due to its lack of existence or test agency policy preventing disclosure). They also argue that test producers have lobbied against federal funding for test research studies and that they only release data to the "right types of researchers."

Opponents argue that test agencies already provide a great deal of data to researchers and that they continuously release reports, present papers, and publish results of research studies on tests and testing. They argue that the current form of the legislation would place an undue burden on test agencies by requiring them to provide the government with any study based on data they provide. This would result in the test agency either reducing their data access or placing certain restrictions on independent researchers. They also claim that the legislation is vague with regard to what constitutes "a study, evaluation, or statistical report." It is common that preliminary draft reports are not disclosed because revisions may be necessary following a review period. Premature release of statistical reports or research studies may be used to mislead the public and policy makers.

CHARGE: The commission/study could provide the Committee with an assessment of the "state of the art" with regard to testing research. Such an assessment could include an analysis of available data for such studies, the current knowledge and information limitations regarding testing research, and the availability of federal and non-federal support for such research. The analysis could also include an assessment of the research and evaluation components of the legislation to determine its enforceability and reasonablenesss of the study requirements.

SUMMARY

The most salient opposing themes of the proponents and opponents of the proposed legislation can be summarized as follows:

PRO	CON
The legislation is a reasonable federal effort to assure due process for individuals, further the public interest, and increase the understanding of admissions tests and their use. Admissions testing is dominated by a few large organizations which have significant influence over individuals, academic institutions, and occupational organizations; therefore, testing agencies should be subject to close public and professional scrutiny (in a manner similar to that required of public utilities).	The legislation is a first step toward improper federal regulation of testing and admissions processes. Many of the charges leveled against testing are, in effect, attempts to "shoot the messenger." This legislation could damage a valuable public service which voluntarily does much of what the legislation would mandate. The form of the proposed legislation is often vague, occassionally unreasonable and, in some instances, violates existing legal protections of privacy, copyright, and academic freedom.

The debate and discussion stimulated by these two bills has done much to clarify the issues and problems associated with federal legislation in the admissions testing arena. The testimony has made a significant point that we share. That is, the consequences of the central features of the legislation are not well understood. Test item security has been a tenant of the testing process. Any definitive statement on the effects of test item disclosure would be both misleading and premature. The Committee faces four fundamental questions as they act on these bills:

1. Is the legislation valid? That is, will the provisions of the proposed legislation accomplish the purposes which they set forth?

2. Can the legislation be improved with further understanding and knowledge gained from existing or recommended studies?

3. Is such legislation necessary?

4. If so, is the federal government the appropriate body to enact, administer, and enforce such legislation?

Not all of these questions can be answered by the four phenomena noted in our statement:

--The report of the Committee on Ability Testing of the NAS
--The analysis of the recent New York law and its consequences
--A review of the current and revised Standards for Educational and Psychological Tests
--A congressionally established study commission or mandated NIE study on admissions testing.

It is our belief; however, that these events, studies, and reports will significantly assist the members of the Committee in making their decisions on the proposed legislation. We have already provided the Committee with the names of a number of individuals who, from a basis of research and scholarship, are able to provide expert testimony on the proposed legislation. We will be pleased to further assist the members and staff of the Committee as they continue their discussion on these bills.

THE AMERICAN EDUCATIONAL RESEARCH ASSOCIATION

contact:
David H. Florio, Director
Governmental and Professional Liaison

National Academy of Sciences - National Research Council
Assembly of Behavioral and Social Sciences

ATTACHMENT -A-

COMMITTEE ON ABILITY TESTING

*Wendell R. Garner (Chairman)
Dean of the Graduate School
Yale University
(psychology)

Marcus Alexis
Department of Economics
Northwestern University

William Bevan
Provost
Duke University
(psychology)

*Lee J. Cronbach
School of Education
Stanford University
(educational psychology,
psychometrics)

*Zvi Griliches
Department of Economics
Harvard University

Oscar Handlin
History Department
Harvard University

Delmos Jones
Anthropology Department
City University of New York

Lyle V. Jones
Dean of the Graduate School
University of North Carolina
(psychometrics, statistics)

Philip P. Kurland
School of Law
University of Chicago

Burke Marshall
School of Law
Yale University

William James McGill
President
Columbia University
(psychology)

Melvin R. Novick
Lindquist Center for Measurement
University of Iowa
(statistics, psychometrics)

Lauren B. Resnick
Learning Research and Development Center
University of Pittsburgh
(psychology, educational psychology)

Alice Rossi
Department of Sociology
University of Massachusetts

*William Sewell
Department of Sociology
University of Wisconsin-Madison

Janet T. Spence
Department of Psychology
University of Texas

Alan A. Stone
School of Law
Harvard University
(psychiatry, law)

Mary L. Tenopyr
Human Resources Laboratory
American Telephone and Telegraph Company
(industrial psychology)

*John W. Tukey
Bell Telephone Laboratories, Inc.
(statistics)

E. Belvin Williams
Educational Testing Service
(psychology)

Edwin P. Hollander, Study Director
Alexandra K. Wigdor, Associate Study Director
Susan W. Sherman, Senior Research Associate
Gladys R. Bostick, Administrative Secretary

*Member of the National Academy of Sciences

March 1979

3/79

PREPARED STATEMENT OF DONALD G. ZYTOWSKI, ED. D., COUNSELING PSYCHOLOGIST AND PROFESSOR OF PSYCHOLOGY, IOWA STATE UNIVERSITY, ON BEHALF OF THE AMERICAN PERSONNEL AND GUIDANCE ASSOCIATION

I have been asked by the officers of the American Personnel and Guidance Association to prepare a statement on the two bills, HR 3564, introduced by Representative Gibbons, and HR 4949, introduced by Representative Weiss and others.

I have been a member of this organization since I was in graduate school preparing to become a professional counselor and am greatly honored to be asked to represent its more than 40,000 members.

I have been for the last 14 years a counseling psychologist on the staff of the Student Counseling Service at Iowa State University. In this capacity, I see numerous students each week, and endeavor to use my professional skills to help them with their educational, career, and personal concerns. I am also Professor of Psychology, teaching counselors-in-training the skills and competencies of the profession. As well, this year I am serving as the president of the Association for Measurement and Evaluation in Guidance, a division of the Association, and have just retired as editor of the division's journal, Measurement and Evaluation in Guidance. I am also a Fellow of the American Psychological Association, a licensed psychologist in the State of Iowa, and a diplomate of the American Board of Professional Psychology. I would emphasize my very first stated qualification for this task: that I see college students, on a day-by-day basis, about their very serious and real concerns in attaining their educational and life goals.

No one can disagree with the stated purposes of HR 3564 and 4949: to ensure equal rights and fair treatment of test takers, to promote more knowledge of tests and admissions practices, and to improve the quality of standardized tests. Professional counselors are well aware of the anguish of some students as they take admissions tests and many have seen instances of the misuse of test scores in the admissions and placement process. I endorse these purposes without reservation.

A general reading of the bills gives rise to the implication that there are a number of specific grievances on the part of applicants to colleges and post-graduate programs which must be righted: exclusive reliance on a test score, profiteering by the test developers, secret reporting codes, serious errors in computing scores, applications with noticeably low validity. While testimony accompanying early hearings on these bills recounted such charges, from my perspective as a day-to-day counselor, and as one who keeps abreast of the publications relating to my profession, I know of no credible reports of such abuse as widespread. After reading the provisions of these bills, I conclude that they will work far more damage and hardship than that which they are intended to correct.

Mr. Gibbons' Bill, HR 3564, appears to call for two things: more information about the nature and purposes of the tests, and for certain requirements in the use of occupational admissions tests. Many tests in current use, such as the

Graduate Record Examination or the ACT Assessment, appear to me to exceed the information requirements of this bill already. The ACT test does an excellent job of presenting results for the understanding of the test taker, and the GRE goes to the trouble of including in its information booklet an entire test, with a correct answer key, and norming information. I believe that other test producers could be induced to come up to this standard without taking the time and attention of the Congress of the United States. Concerning the occupational features in Mr. Gibbons' bill, it seems to me that its requirements are amply provided for by the EEOC regulations on employment testing. Is it necessary to duplicate this already functional law?

Mr. Weiss' Bill, HR 4949, is much more specific in its requirements and merits more extensive comment. I will limit mine to several major features of the bill, since I have read testimony from test publishers which address many of the technical and legal specifics which concern them most directly.

Much of the bill concerns itself with the misuses of tests. Yet, the proposed regulations are applied to the producers of tests instead of the users. Of course, a test publisher should reasonably be expected to offer evidence of the quality of any test. But the responsibility for showing the validity of a test in any specific application should rest with the user. It is this principle which is embodied in the EEOC regulations on employment testing. If a low predictive utility is revealed in a given application, it behooves the user to select a better test or other indicator, or to examine the nature of their criterion (usually first year grades) to determine whether it is contributing to the problem. Many test producers provide services to compute these coefficients on a custom basis; perhaps the National Institute of Education could be asked to provide assistance for institutions which cannot avail themselves of such services.

The bill does not acknowledge another use of tests, which I believe might be interfered with if enacted. In many state colleges and universities, admission is mandated by state legislatures, often on the basis of rank in one's high school graduating class. Tests which are used for admissions purposes in other institutions are used to assist in placing the new student in courses where he or she may have a reasonable expectation of success. Validity information in this application is much more difficult to obtain owing to the smaller numbers and lack of correspondence between the many courses which are affected.

It is hard not to conclude that the bill seeks to discourage the use of tests. While it is stated that the purpose is to encourage the use of multiple predictors, the bill only asks for disclosure about one such predictor--tests--and fails to address other extensively used predictors, such as previous grades or rank in class, or letters of recommendation. Should it not ask for the reliabilities and validities of these predictors to be made public as well?

The New York "Truth in Testing" law recently enacted may be instructive to the framers of HR 4949. It has been reported that the sponsors of the medical and dental college admissions tests will not administer these tests in New York. On the basis of what information will New York's students be admitted to medical and dental schools? Grade average? There are many reports of "grade inflation," which would reduce the value of grades in selecting students for post-baccalaureate education. Recommendations? Personnel psychology has published many reports of research which show that letters of recommendation are the least effective predictors of performance. I am impressed by the irony of the situation. Tests were first introduced to correct unfairness in access to opportunity: the "Truth in Testing" laws could very well return us to that old wrong.

Although Mr. Weiss' Bill names governmental agencies in his definition of test score recipients, I note that the government is absent in the definition of test agency. Governmental units of all kinds are among the highest volume test producers: State, county, and municipal civil service tests, the PACE Exam of the Federal Civil Service (which in my experience produces as much anxiety among college seniors as the SAT does among high school seniors), and the Armed Services Vocational Aptitude Battery (whose informational materials appear to me to not meet the standards of the bill). All of these organizations provide post-secondary education. In fact, the appeal of the Armed Services is to continue one's education. If the bill is to achieve its stated purposes, these tests should surely be covered by its provisions.

Further, do the bill's sponsors realize that some states are both producers and users of college admissions test. Can a federal law apply to these circumstances? Should a high school student be protected when she or he takes the SAT, but not when he or she takes the state-developed test?

The provision of Mr. Weiss' Bill which is most controversial, clearly, is the disclosure of test items with their correct answers. Such a practice seems quite consistent with the growing openness seen in the Freedom of Information Act, open meeting laws, the Truth in Lending Act, and the like. Presumably for tests, this provision would allow the test taker to ascertain whether any error has been made in computing or reporting his or her score. In contrast with the information revealed in the foregoing laws, any test item with its correct answer which is made public becomes useless and must be replaced in the next administration. As I am sure the test producers have testified, and as even the rawest elementary school teacher can verify, it is a difficult process to produce a good, fair, and valid test. What is more, the revelation of test items will without doubt lead to practices which will reduce the range of scores earned, which in turn will reduce the validity of the test.

As an alternative, if a test taker suspects a serious error in her or his score, could not a person be designated in a school who would receive the test taker's answer sheet and scoring key in order to do the manual check of the machine score? That person could even counsel with the test taker about the nature of their wrong responses, if the test were scored correctly. What better person could be designated than the school's counselor, who very likely had the responsibility for administering the test in the first place, and who saw the test items but whose regard for professionalism does no damage to the future of the test by revealing them. This same counselor probably also has the expertise to insure that the augmented information about the test could be used to reduce the test taker's anxiety, or could apply anxiety-specific counseling as well.

If the purpose of this provision is educational as well, publication of a copy of the test, as several publishers now do and as the recent California law specifies, should suffice without the potential for damage that revelation has.

Several bodies are attempting to prepare more extensive responses to the proposed legislation. National Science Foundation, the National Consortium on Testing of the Huron Institute, and the Association of Measurement and Evaluation in Guidance are three that I know of at this writing. Would it not be more reasoned to postpone action until these groups, representing producers and users as well as personally disinterested parties have had an opportunity to produce their analyses?

In summary, while I know of misuses of admissions testing, I do not believe that they are so widespread nor so damaging as to require legislation as comprehensive as HR 3564 and HR 4949. I would rather that such legislation be based on the findings of widely published and verifiable information, such as may be forthcoming in the near future. If the Congress is intent upon enacting such a law, I would respectfully request that it consider incorporating the following modifications:

1. Require that the reliabilities and validities of all variables used in admissions decisions be calculated, and available upon request from the user, such as the EEOC guidelines on testing in employment require.

2. Be certain that all tests which are used in admission to educational opportunities, particularly those promulgated by governmental units, are included in the coverage of the law.

3. Encourage the publication of a sample test, with scoring key and norms, in the information about the test which is given to test takers before they actually take the test.

4. Permit a disinterested and professional third party--a counselor--to examine the test taker's answer sheet, at their request, for scoring errors.

PREPARED STATEMENT OF DR. DAVID G. STRATMAN, DIRECTOR, CAMPAIGN FOR THE FUTURE

Mr. Chairman and Members of the Subcommittee:

My name is Doctor Harold E. Jarvey, Jr. I am Executive Director-Secretary of the Federation of State Medical Boards of the United States, Incorporated, which was founded in 1912. The membership consists of the medical licensing boards of the fifty states and the District of Columbia, together with Puerto Rico, Guam, Canal Zone and the Virgin Islands. In addition, eight independent state osteopathic examining boards are members and the ten Canadian Provincial Licensing boards are affiliate members.

Mr. Chairman, although Henry Ford, an acknowledged genius observed that "History is bunk", I would respectively submit that for ordinary mortals he who does not know history is condemed to repeat the mistakes of the past. The history of medical licensure is a case in point. Until 1968 each state board developed and scored its medical licensing examination. The system was chaotic and state board examinations lacked uniformity. In the majority of states exams were largely subjective. The test methodology was relatively unsophisticated. It would be difficult, if not impossible, for two or more examiners, grading the same exam, to arrive at the same scores. Reliability and validity were terms foreign to the vast majority of board members.

Over the past fifty-six years, thousands of dollars and countless man hours have been spent by dedicated, intelligent individuals to create a high quality examination which all states would accept. Through the efforts of many, this goal was finally achieved in 1968 and "FLEX" was born. The acronym stands for Federation Licensing Examination. The "x" was added to make it more euphonistic; otherwise, it would be the "FLE" examination. Not very dignified!

1066

I could explain how "FLEX" is constructed. How the questions are evaluated for reliability and how the scoring is done. These issues have been dealt with in great detail by others and there is no need for me to repeat.

One of the problems that this proposed Legislation intends to correct and, I quote, "Testing of skills for entry into an occupation whether of a profession, craft or trade nature, is a critical factor governing the even flow of individual skills in interstate commerce and seriously affects the nation's capability for economic growth". I would agree. What has "FLEX" accomplished in resolving this problem?

Eight states participated in the FLEX program in 1968. In the relatively short span of ten years, with Florida entering the program this year, all 50 states, the District of Columbia, Canal Zone, the Virgin Islands, Puerto Rico, Guam, and the Province of Saskatchewan, use "FLEX". This means there is a single examination used across the U.S. and its territories. The Federation is justly proud of this achievement. This not only gives a sense of confidence in the competence of physicians entering the practice of medicine, but removes barriers which existed prior to 1968. In the years before, states would recognize the examination of certain states as being equivalent to their own, but would not accept others. Reciprocity contracts were entered into by states. These, in effect, said, "You take my licensee and I'll take yours". There was no consistant rationale as to how these decisions were made. As a result, physicians could not move from one state to another without being subjected to another examination. This was a "critical factor governing the even flow of individual skills in interstate commerce". This problem does not exist in 1979. With the universal acceptance of "FLEX" there is increased physician mobility.

How will the proposed Legislation affect the remarkable progress that has been made in medical licensure with the advent of "FLEX"?

As others have testified, the pool of questions is limited. The requirement that the candidate can request and must be given a copy of the exam with all the correct answers and those missed would, within a short period of time make it a farce. By reviewing several exams, any reasonably intelligent, highly motivated individual would be able to make a passing score. The validity and reliability of questions and security of the entire exam would be non-existent.

The proposed requirements for fee disclosure and justification are also inappropriate in the context of the state medical boards. Obviously, the examination fees levied by such boards are already a matter of public record. The justification for those fees is readily apparent. As arms of the various state governments, these boards are charged with maintaining the qualification of new physicians - a role traditionaly accepted by the states. By virtue of that status it would be inappropriate for the Congress to attempt to regulate, or oversee their activities. This is particularly the case since there are no alleged abuses by the state boards that would call for reform.

From the standpoint of the Federation of State Medical Boards, the overriding concern is, "Does this committee, through this proposed Legislation, wish to throw out the baby with the bath water?" For the reasons stated, as well as others, I am reasonably confident, if these bills were passed, some of the state medical boards would withdraw from the "FLEX" program. Such boards would once again develop their own examinations, with all the inequities and irrationality that you wish to correct. All of the efforts expended and progress made might well cease. Medical licensure could well revert back to the pre 1968 days. I would think the same concerns would be voiced by such groups as Dentistry, Law, Nursing, Osteopathy and Pharmacy.

The Federation opposes HR 3564 and 4949 and urges that they not be passed in their present form. There has been no groundswell for the Congress to become involved in medical licensure. In my judgment the stated meritious objectives will not be achieved. Progress will not result.

I thank you, Mr. Chairman, for the opportunity of testifying. I will be happy to answer any questions you or the committee members may have.

Prepared Statement of Dr. Harold E. Jervey, Jr., Executive Director-Secretary, Federation of State Medical Boards of the United States, Inc.

Mr. Chairman:

I am David Stratman, Director of the Campaign for the Future, an organization committed to the revitalization of public education. I wish to thank you for the opportunity to offer testimony on H.R. 4949, the Educational Testing Act of 1979.

We welcome and endorse the Educational Testing Act as an important step in providing to students and parents information about a process which has immense impact on students' education and on the opportunities which will be open to them upon graduation. We see the importance of this legislation to be not only what its enactment will accomplish; we view it as the opening round in what must become a national debate on standardized testing: on its uses, its limitations, and on its effect on the educational process. Few issues strike so deeply to the heart of educational policy and practice as the issues raised by standardized testing.

The concern of Campaign for the Future is with standardized testing at all levels of education. H.R. 4949 speaks only to standardized testing at the college level and beyond. I would like to comment on the specific issues raised by the Educational Testing Act, and also to extend the discussion to questions related to standardized testing which pertain as well to elementary and secondary education.

The most critical question about any form of testing is, "What do these tests measure?" The question is important because it brings to attention not only the educational usefulness of the test, but also raises questions involved with test design and construction. Mr. Robert Solomon, Executive Vice President of Educational Testing Service, raised this question in his testimony before this Subcommittee on July 31. The question involved is one of cultural bias:

> ...the argument is made that tests of academic ability are biased because they represent middle class culture. These tests do reflect skills and knowledge considered important in the mainstream culture, in many jobs and in higher education. But the fact that a test mirrors the common culture is a poor basis for calling the test biased.

The case for removing the shroud of secrecy from educational testing does not hinge upon the charge of cultural bias in testing. Students and parents have a right to know on what basis their lives are being directed toward one option or another apart from the general validity of the bases of these judgments.

Nevertheless, the claim that standardized tests simply "mirror the common culture" is not true in the sense in which it is intended in the statement. The "cultural bias" inherent in standardized testing is not a peripheral problem, to be resolved by a simple modification of particular test questions. It is inherent in the design and the purposes of the tests.

The design of a test is a function of the use for which it is intended. If a teacher wants to know how well her students have learned their multiplication tables, she will ask her students some of the problems they have studied. That is called a criterion referenced test, and is used every day by classroom teachers. It is a useful kind of testing, because it helps the teacher design the next step in the students' instructional program. This kind of test is not inherently biased.

If a teacher has been told to grade her students on a "curve" with some A's, some F's, and some students falling in between, she will need to design a test which will procude that distribution of scores. A test designed to produce a "normal" or "bell-shaped curve" is called a norm-referenced test. Its purpose is not to measure what students know, but to sort them out along a bell-shaped curve. If all of a teacher's students have learned their multiplication tables well, so that they would get all A's or B's if tested on their tables, the teacher would have to design a test that would sort the students out on some other basis than their knowledge of their times tables.

This could be done in any number of ways---by presenting the questions in confusing language that only some of the children would understand; by making the test a speed test that some of the students would not respond to as well; or by testing these students on some different material and then claiming that the children were being tested on their "mathematical ability" or "mathematics aptitude."

Standardized norm-referenced tests use all of these techniques to produce test scores that the test-makers claim reflect students' "mathematical ability" or "verbal ability." The standard achievement tests frequently administered to grade- school youngsters are tests of this type.

Tests of this type are inherently biased, because they are designed to produce a predetermined distribution of scores irrespective of the students' knowledge of what they have been taught. To design a norm-referenced test, the test-maker must first decide which students exhibit what he considers to be intelligence or achievement. He then designs a set of questions that fulfill his expectations of who is smart. If students whom he has selected perform poorly on a certain question, he rejects that question and selects a new one. And vice-versa: if children whom he has judged to be poor students perform well on a question, that question is rejected.

Questions like "What is intelligence?" or "Who is smart?" or "What is student achievement?" are questions that the test-maker answers on the basis of his assumptions about students before he constructs the test designed to answer these questions.

Scholastic Aptitude Tests and other major standardized college entrance examinations are referred to by their publishers as "performance tests." While they differ in some respects from standardized achievement tests, they are similar in that they are designed to sort students out rather than to test what they know.

In the case of SAT's, the tests are designed to sort students in such a way as to predict their success in college. This is their value to the colleges which subscribe to the Educational Testing Services services. Questions are selected for the tests not because they reflect some curriculum which all the test takers have shared and thus can be expected to know. They are selected because those students who tend to do well in college tend to answer them correctly. If students who have not historically done well in college answer certain questions correctly, but students who have done well do not, those questions are rejected in favor of better predictors.

Thus standardized entrance examinations are not "mirrors of society" in the sense that they reflect the skills and knowledge of students from different backgrounds. They are mirrors only in the sense that they reflect and enforce historical patterns of social hierarchy. They are not mirrors of achievement and ability but gatekeepers on the paths to social status. Class and race bias are inevitably part of their design, because these factors cannot be disassociated from the predictive purpose of the tests.

A critical question related to standardized testing is, "What is the educational effect of these kinds of testing?" This question emerges most forcefully in considering norm-referenced testing, especially at the elementary and secondary level.

Standardized tests are used to make decisions about students at many critical junctures in their education. They may lead a child to be labled or "tracked" as having only a certain level of ability. They may influence guidance counselors and teachers to advise or treat a child in a certain way. They can affect a student's chances for college or scholarship money. Their impact and potential for harm on students' lives are immense.

But the sorting process and the bell-shaped curve of the test can have another effect that is of a more day-to-day and finally more important nature. Standardized tests are an outside "authority" that seems to be coming from on high. Parents and teachers may feel uneasy about these tests and the judgements which they impose, but still find it difficult to challenge them.

Because the tests compel and legitimatize the sorting and labelling of the students, they necessarily have an enormous effect on how teachers and parents view the students, and how the students view themselves.

One half of students tested overall, no matter how well they have mastered the material, are told that they are "below average." Each student is told by the test that his performance is important only to the extent that he compares against other students taking the test--since that is all these tests can do. (Remember that the performance is based on "aptitude," not knowledge of the curriculum.)

Standardized intelligence and achievement tests, in other words, bring into the classroom and into the lives of the students a view of themselves, of education, and of each other that can only damage their educational development.

A third area of our concern is with the barrier of secrecy which has been erected around testing. H.R. 4949 speaks pointedly to this issue.

From early childhood through entry into professional schools or jobs, standardized tests make significant judgments about our children. Yet no portion of the public school system is less understood than the facts about standardized tests, their assets and their limitations. Campaign for the Future is deeply concerned that testing programs are being authorized, paid for, instituted or withdrawn and evaluated without any broad public discussion on many of the momentous issues raised by these programs. As parents and teachers, we feel it is important to achieve a greater degree of accountability in the testing process, yet such a review or public discussion is difficult or impossible because of the intense secrecy that surrounds these tests.

Due to the mask of "security," the public has been unable to evaluate dispassionately the claims of test makers, the uses by public school systems and institutes of higher learning, or the many reported abuses. Information about the tests and the testing process is minimal and is rarely shared "after the fact" with those most intimately concerned students and their families. Yet these tests serve as gatekeepers to the future for our children: they develop patterns of expectation by teachers and school systems that often are totally unrelated to a child's true capabilities; and because the power to test is the power to influence curriculum, they tend to reinforce curriculum patterns that respond only to the ultimate goal of "looking good" on the national tests.

While H.R. 4949 would by no means address these critical questions directly, by lifting the shroud of secrecy from these tests, it would for the first time allow a broad and informed debate over their construction and uses. The debate would allow both the public and the individual student to put the tests in a more reasonable perspective.

Campaign for the Future has a further concern with standardized testing which relates to its effect on the future of education in the United States. Many experts on educational testing have pointed to the history of standardized testing as a device for social control.

Its first uses, both abroad and in the United States, have been as a means of sorting out those individuals with backgrounds and values most like the most powerful groups in society. The "gateKeeper" function of testing is to sort out those who will "succeed" in these terms, and to justify the rejection of those who do not. As it was with immigrants in the 1910's and 1920's, so it is now with the children of the working class: standardized testing is constructed to justify the advancement of the children of the middle class, and to assign to working class children a low statusand a diminished future. The tests justify in terms of "aptitude," "ability, " or even "intelligence," what is in fact a clash of values and family background. The tests are thus used to justify an unequal society which promises in the future to become more unequal.

It is this historical pattern which concerns us most. The use of standardized testing has been greatly increased in the past several years, at the elementary and secondary level, and promises to be even more influential in children's lives in the future. As economic opportunities diminish in what will likely be a contracting social hierarchy in the next decade, there will be powerful pressures on the educational system at all levels not to educate students but to sort them out, and to justify their selection with supposedly "objective" measures. This eventuality would be disastrous for America's children and for the schools; indeed, it would mean a historic reversal in the direction of American society. We see the pressures for increased testing to be akin to the pressures for tuition tax credits and other attempts to make education the privilege of a few. Their aim is to lower the society's expectations of what its children can achieve.

I would like to make one further comment in support of this legislation. Critics of the bill have alleged that H.R.4949 is a measure for further federal control of education. In fact, it is nothing of the kind. What exists now in education is a huge testing establishment, with immense effect on millions of students' lives, which is veiled in secrecy and accountable to no one. This legislation does not propose government regulation of the industry; it merely would enforce the "deregulation" by the industry of its own processes, by making them open to test subjects. The bill does not interrupt the rights of private individuals: it restores them.

We urge this subcommittee to act favorably on this legislation. Thank you, Mr. Chairman.

PREPARED STATEMENT OF THE NATIONAL LEAGUE FOR NURSING, NEW YORK, N.Y.

STATEMENT FOR THE RECORD OF HEARINGS

ON H.R. 4949 AND H.R. 3569

The National League for Nursing, a coalition of nurses, other health professionals, consumers and health agencies -- 17,000 individual members and 1,800 agencies -- all of whom work together to help meet community nursing needs and to assure high quality nursing education, respectfully submits this statement for the Record of Hearings on H.R. 4949 and H.R. 3569.

NLN provides preadmission and achievement test services to state-approved schools preparing registered and practical nurses. The League also serves as the test service agency for the State Board Test Pool Examination for both registered and practical nurse licensure. Each of the licensing examinations is a nationally standardized test, administered twice a year on uniform testing dates.

The National League for Nursing has had many years of experience in admission testing in the areas of guidance and placement of students in schools of nursing. The first edition of the Pre-Nursing and Guidance Examination (PNG) for use by registered nursing programs was published in 1946 and an edition of this battery has been continuously available for the last 33 years. NLN has provided a preadmission test battery for practical or vocational nursing programs since the 1950s. Although we understand the serious concerns of many consumer groups and students over a procedure that plays such an important role in the education and careers of many, the League believes that there are serious flaws in some of the provisions of the proposed legislation. Therefore, we wish to focus our comments on the issue of disclosure of test content.

Legislation advocating disclosure of test questions and answers seems hasty, especially considering that the result might be the drastic reduction in standardized testing available for guidance and placement of applicants to post secondary programs and the subsequent use of the subjective criteria in many cases for this purpose.

It should be remembered that fairness to the examinee is the primary objective of test security. The practice of treating test content as secure material evolved so that no test taker would enjoy an advantage over another through prior familiarity with the specific questions included in the test. While granting all examinees access to test content before testing may seem to be providing fair and equitable treatment to test takers as well, realistically this is not so. Given the wide diversity in circumstances for examinees, factors - not the least of which is economic - will decrease the opportunity of getting the questions for many.

Keeping test forms secure ensures that each test taker is being presented with the same test problems. Standardizing testing conditions are necessary for sound test development and statistical studies for equating scores from different editions of a particular test. The customary equating procedures require the comparison of student performance on a new set of questions with performance on a set of questions previously administered within the testing program. One widely used method of carrying out this comparison is to administer, as part of a new test, a sample of questions that had been previously administered as part of the same program.

Only when all test takers encounter testing materials for the first time under similar examination conditions is it reasonable to use performance on the larger body of material from which the test is drawn. If all such previously administered questions had already been released and widely circulated among examinees, alterations would be required in the equating process since it would no longer be possible to treat both sets of questions as identical for purposes of establishing the relationship between new questions and the previously administered questions.

A standardized test is valuable particularly in two kinds of situations: situations in which comparisons have to be made; in which there are large numbers of people about whom decisions need to be made, but for whom the decision-maker has no common or comparable data. There are a great number of situations in education where comparisons need to be made. One important example is guidance and placement of applicants to educational programs whether the concern is to select the best qualified for a limited number of available openings or to allocate teachers and facilities to meet the learning needs of prospective students in the country.

Standardized tests are also very valuable as predictors of future behaviors. As part of NLN's service to test users, the League has conducted studies on our admissions tests to determine their usefulness as predictors of performance in schools of nursing. The most recent validation study of the current edition of Pre-Nursing and Guidance Examination affirmed the usefulness of that instrument as an aid in making admissions and placement decisions for schools preparing registered nurses. The results indicate that the Pre-Nursing and Guidance Examination battery functions well as an indicator of future performance in schools of nursing, and on the licensure examination. Although each test in the battery contributes to prediction of success within the nursing program, conduct of the study enabled NLN to identify those tests which contribute the most.

The construction of sound standardized tests is a long arduous process and the investment of money and effort is a formidable one. The first step in constructing a standardized test is development of the plan and specifications for the test. Then items have to be written to fit the specifications, reviewed, edited, and tried out on groups similar to those for whom the test is intended. Next, item analysis must be done. On the basis of the results of item analysis, the items for the final form of the test are selected, or items may be revised and new items written and tried out again. When all the items for the final form of the test are ready, the tests have to be printed, a representative national sample must be recruited, and the tests administered to the sample of students. Tests must then be scored, raw scores tabulated, and norms developed. At this point additional data need to be gathered and analyses carried out to permit determination of reliability and correlations with other measures. Manuals for administering the test and interpreting the results must be prepared.

This process is obviously a very painstaking, time consuming and costly one. Test disclosure would require that portions of this process be eliminated or modified drastically because of time constraints and that other portions be repeated each time a test is administered. The most obvious increase in time and expense will result from the additional test items that will need to be written and readied for use. Witness the recent example of New York State where as a result of the passage of the Admissions Testing Act of 1979, some testing programs are charging as much as $5 more than students elsewhere to take exams. The National League for Nursing would accrue large cost increases with the construction of new tests and associated printing costs. There also exists the very real danger that there will be a depletion of the pool of high quality items. Especially where much of the information to be tested is of a discrete, technical nature, as in the health professions, there is a limit to the number of questions that can be developed.

Another very critical point that needs to be more closely examined is whether the disclosure provision of the legislation will be able to accomplish what it is intended to accomplish. That is, releasing the test questions alone would not be of much benefit to the student without any knowledge of other factors involved in developing the items such as the relationship of individual questions to the test as a whole or the level of difficulty of the items. Understanding this requires specialized training . If the intent of the legislation is to reveal the questions so students will be able to identify the areas where they are academically weak, then review books are available for all the major admissions tests and could surely be used for that purpose. If test items are required to be disclosed, they will simply provide an additional sampling of test questions, some of which are always found in review books now. In addition, most test administrators already inform applicants of the general areas to be covered and after testing, tell him how they did in each specific subject area. The State Board test for registered nurse licensure most assuredly does this.

Proponents of the legislation also believe that the public release of tests and answers could help to detect cultural bias in the test items. The exact opposite situation may well occur considering that standardized testing provides the data for research on cultural comparisons in test performance. There is the danger, considering a greater availability of transportation, time and financial resources to white middle class students, that they would be more likely to acquire test items and gain more understanding of the entire testing procedure because of having more support systems available to them. Test users would be required to obtain personal data from examinees being tested in order to comply with the terms of the proposed law. Conceivably, some test subjects might object to providing personal information - such as family income - in the testing situation.

At the present time there are many serious repercussions occurring in New York State as a result of the State's new Admission Testing Law, requiring that publishers of standardized admissions tests file test questions and correct answers with the New York Commissioner of Education. The State Education Department has estimated that 20 of the 26 college and graduate school admission tests covered by the law would be withdrawn from the State after January because of the cost of developing additional questions. In addition, the frequency of test sessions will be reduced, making it more difficult for examinees to select a convenient time for testing. Proposed legislation would make it more difficult for NLN to accomodate the needs of handicapped students. Before enacting legislation on testing of this scope, legislators would be well advised to take the time to examine the implications and repercussions of "truth in testing" in New York State.

PREPARED STATEMENT OF THE AMERICAN NURSES' ASSOCIATION

As the national professional membership organization for registered nurses, the American Nurses' Association is dedicated to improving the quality and availability of health care services for all Americans. A key factor in meeting these goals is an adequate supply of competent nursing manpower. There is opportunity and need for developing and utilizing talent from all sectors of our society, most certainly including minorities and educationally disadvantaged groups.

Standardized testing performs a critical role in cultivating this manpower resource. Tests are used to identify students of promise to provide feedback for continuing development in both educational and professional practice settings, in licensure to determine ability for safe and effective nursing practice, and in certification to assess qualifications in specialized fields of nursing. Through national standardization, tests promote educational opportunity and mobility of professional workers. While such tests should always be used in conjunction with other indicators of abilities and qualifications, we know of no fully adequate substitute assessment method.

The American Nurses' Association therefore wishes to insure that aspirants to and members of the profession have timely access to highly valid and fair testing programs without artificial barriers to candidate success. We do not believe that the proposed legislation will contribute to this purpose. In previous testimony, others have raised serious questions about the effects of such law, including concerns about the prospective negative impact on minorities, quality of tests, accessibility and costs of tests to candidates, and misdirection of student learning. We believe any case for regulating testing should be based upon a documented need and a thorough analysis of costs and benefits to be achieved.

We ask that this statement be made a part of the permanent hearing record.

PREPARED STATEMENT OF ROBERT GLASER, PRESIDENT, DIVISION OF EVALUATION AND MEASUREMENT, AMERICAN PSYCHOLOGICAL ASSOCIATION

The Division of Evaluation and Measurement of the American Psychological Association supports the intent of both H.R. 3564, the Truth in Testing Act of 1979, and H.R. 4949, the Educational Testing Act of 1979, to improve the quality and availability of information about tests and testing. However, we do oppose certain aspects of both bills because technical considerations would work to vitiate the fulfillment of this intent.

We support the following: First, that students and other persons who take tests should have a sample of test items available to them before the test is given so that they can see the kind of items they will be confronted with in the actual test. Second, that students also need some assurance that their answer sheets have been correctly scored and recorded. Third, that they need to know in general about the relationship of test scores to the achievement or performance being predicted. Finally, that the consumer needs to know that the items are fair and of good quality.

1084

We believe, however, that none of these purposes is served by Section 5 of H.R. 4949 requiring public disclosure of all test items. In fact, disclosure of items used on standardized tests has many disadvantages (most of which militate against the students themselves) which far outweigh any possible advantages. In some cases, a complete sample test with instructions and explanations of correct answers is provided to students who sign up to take the test and study guides frequently are available for many other tests.

Accurate scoring of test results has not been a serious problem with modern, standardized tests. The number of checks and balances in computer-scored tests makes errors extremely unlikely. Furthermore, it is and always has been standard practice to recheck the scoring on a test when doubts or disputes arise.

Extensive statistical data are available on the relationship between test scores and the achievement that they are intended to predict. This information has been available to anyone interested from numerous and varied sources including publications by test developers, college publications, and the technical journal literature.

Regarding the consumers need to know that test items are fair and of good quality, we believe that item disclosure will work against the construction and use of such items. Quality items on standardized tests take months and even years to develop in order to insure that they are fair, that they validly predict future performance, and that they are unambiguous. To disclose such items after every administration of a test eventually could result in the hasty construction of new items which, in all probability, would be less fair, less valid, more ambiguous, and generally of poorer overall quality. Moreover, the effect of disclosing items inevitably would result in fewer services provided to the student. For example, the Graduate Record Examination now is given six times a year, in addition to special administrations for handicapped persons and for persons whose religious convictions prevent their taking tests on Saturdays. Such special services to students could be seriously curtailed.

Additionally, we wish to point out that decisions currently requireing test scores as one source of information will continue to be made, and if tests become less reliable and less valid, then more subjective judgments of student achievement and qualifications will be relied upon. It is also important to realize that release of test items will not assuage the inequities engendered by the unequal quality of education.

Most of the provisions of these bills appear to be in the public interest, and the Division fully supports the movement toward more accountability in testing and more public knowledge regarding how test information is used. At the same time, we ask the Congresspersons involved to consider the deleterious effects that hastily constructed, nonreusable tests will have on all students, majority and minority students alike. We believe that students, themselves would be the principal losers if the provisions of Section 5 of H.R. 4949 were to become law.

We would, of course, welcome the opportunity to aid you in any way possible toward improvements in these bills. Please feel free to contact either of the undersigned to arrange a meeting with representatives of the Division if this would be helpful.

Robert Glaser LE

Robert Glaser, President
Division of Evaluation and Measurement
American Psychological Association

Lorraine Eyde

Lorraine Eyde, Secretary-Treasurer
Division of Evaluation and Measurement
American Psychological Association

PREPARED STATEMENT OF WILLIAM F. JACKSON, D.V.M., PRESIDENT, AMERICAN VETERINARY MEDICAL ASSOCIATION

The American Veterinary Medical Association has reviewed H.R. 3564 and presents the following comments for the record of hearings.

We are persuaded most strongly that testing leading to the licensing and certification of professionals should remain a police power of the respective states, or where appropriate it should be retained by the organized body attempting to raise the competency level of the practitioner of a discipline. We are further persuaded that occupational groups providing testing services to the various state agencies must remain in the capacity of service and advice, with final authority reserved to the appropriate state agency or agencies charged by the laws of that state with the responsibility of assessing the competency of candidates for licensure, certification, or registration.

Testing for competency certification within the various professions (i.e., specialty boards and colleges) should remain a function of the discipline or profession involved until such time as states see fit to require such disciplines or professions to be licensed or certified within the police powers of the states.

We are opposed to any legislation which would in any way require the public disclosure of the examination items. Many valid reasons may be presented in support of an item bank that prepares, validates, stores, uses and analyzes test items that determine the knowledge and ability necessary for certification of an individual to practice a profession. We believe that the elimination of the opportunity to continue this practice would have the following effects.

1. It would markedly increase the cost of creating the examination, and this cost would have to be passed on to the examinee. In small professions such as veterinary medicine and the smaller disciplines within it (e.g. veterinary radiology, etc.), these costs to the individual would be extremely high.

2. With increased costs, the number of times an examination could be administered would be reduced. This would markedly reduce the opportunity now available for multiple state licensing within a reasonable time period (e.g. the first year after graduation). This ability to be licensed in several states presently provides for much greater professional mobility. If the national board examination were given less frequently, this opportunity would be eliminated or greatly reduced.

3. Public disclosure of test items would tend toward the development of new test items which would have a tendency to move toward the trivial, and this would reduce the validity of the examination. The opportunity to test repeatedly for the more critical knowledge and skills permits a more reasonable and practical test which will more nearly measure the knowledge and skill of greatest importance to the public welfare.

4. Public disclosure which would weaken and/or eliminate the test item banks would have a significant impact on small but critical areas of knowledge where only a few items might be appropriate. Public disclosure of such items would permit memorization of pat answers rather than demonstration of valid understanding and knowledge of such material.

5. Public disclosure of test materials would have an inhibiting effect on the development of new testing methods. This is of great concern to the veterinary medical profession which has made significant progress in test design and examination procedures that promise to be of much greater accuracy in measuring candidate ability and skill.

6. Public disclosure of test materials would eliminate the opportunity to use such materials for the certification of candidates on an individual or small group basis since new tests could not be developed for such limited use. Many state veterinary medical practice acts now provide for administering individual examinations under appropriate conditions. We do not wish to deny such individuals this opportunity or be unable to meed such a need.

In summary, we believe that there is no persuasive need for federal legislation in the area of occupational certification and that, in fact, this is an abrogation of state rights. Rather, we believe that legislation written in the perceived spirit and intent of H.R. 3654 would have a markedly disruptive and destructive impact on such testing within and for veterinary medicine. Further, we are of the opinion that the candidates for such examinations are benefited to a greater degree by present practices than would be the case if H.R. 3564 were to be enacted. We request that all those having authority or power to influence such legislation use their position to prevent the enactment of H.R. 3564.

Respectfully submitted,

American Veterinary Medical Association

William F. Jackson

William F. Jackson, D.V.M., President

PREPARED STATEMENT OF ALFRED B. WILLCOX, EXECUTIVE DIRECTOR, THE MATHEMATICAL ASSOCIATION OF AMERICA

Mr. Chairman, I am Alfred B. Willcox, Executive Director of the Mathematical Association of America (MAA). The MAA is an organization of 18,500 members, the largest organization in the world devoted to the improvement of the teaching and content of collegiate mathematics. Our members include persons at all levels and in all branches of the mathematical sciences, but the backbone of our membershp consists of faculty in our nation's 3,500 collegiate and university departments in the mathematical sciences.

The officers of the Mathematical Association of America have recently approved on behalf of all of our members the following statement, which has been sent with minor editorial changes to the members of your committee:

> H.R. 4949
>
> H.R. 4949, the Weiss Bill, calls for full disclosure of all standardized tests - all questions on each test, together with the correct answers.
>
> The bill is introduced in the name of consumerism, but it is certain to lead to increased costs to the consumer and to deterioration in the quality of the tests.
>
> Costs will rise because compromised tests will have to be replaced after each test administration.
>
> Quality will deteriorate because it will no longer be possible to standardize tests. Standardization is accomplished by repeating selected test questions to compare different forms of a test and different test populations, a practice that will clearly not be possible under the disclosure requirement.

Thus the Weiss Bill will lead to more expensive "standardized" tests that have no standards.

For these and other reasons, the officers of the Mathematical Association of America strongly oppose H.R. 4949 and respectfully urge its defeat.

This Statement stands alone, and it summarizes the position of the Mathematical Association of America on H.R. 4949. However, such a brief position statement necessarily omits most of the subtle issues in the proposed legislation. To understand our position fully and to recognize the most serious flaws in H.R. 4949 one must understand these issues.

In this testimony I will amplify the above statement in order to explain more fully why we feel that the Weiss bill would, in the long run, decrease access to education by deserving American youth and be detrimental in many ways to education in the United States.

Standardized testing, in the form we know it in the United States through such testing programs as those of CEEB and ACT, came of age after World War II in response to the drive to open post secondary school education to all qualified students. Colleges and universities suddenly faced serious problems in evaluating the apptitudes and training of vast members of high school graduates vying for a limited number of college openings. Difficult choices had to be made among applicants on the basis of sketchy and unreliable information. High school grades from thousands of high schools spread over the entire geographical and socio-economic spectra were almost useless credentials for deciding which applicants were more qualified and at what level a successful candidate could most efficiently begin studying in a local curriculum. Colleges and universities faced difficult problems in designing their curricula to fit the needs of the American student and in advising schools how to train their students better for entering into the college curriculum.

The opening of post secondary education at all universities to all students, without geographic or socio-economic restrictions, made it absolutely necessary that some norms be established to measure both secondary and post secondary education.

Standardized testing programs were developed to provide the only data for evaluating aptitude and training that could span both populations and time. The information provided by such tests as the SAT and ACT tests provided one data component, to be added to the other components provided by school transcripts, recommendations, interviews, and anecdotal evidence of special talents, that helped colleges select, helped students know where to begin, and helped educational institutions coordinate their curricula.

Without standardized tests, our educational system would quickly drift back toward the elitist European style in which pedigree and position predetermine educational opportunity.

We recognize the dangers inherent in any large scale testing program which is conducted in great secrecy and which can determine or strongly influence future educational and economic opportunities for individuals. However, this description does not fit the standardized testing programs in the United States at all. Both CEEB and ACT have made every effort to disclose as much about their tests and about the grading and scoring procedures as is possible within the minimum constraints of standardization. Colleges and universities have learned to use results of these tests as one data-point among many in making admissions and placement decisions. No institution we know allows a high SAT score to guarantee admission or a low one to deny admission when contrary information is available. And colleges and universities have not diminished their efforts fo find new sources of information about applicants because of the availability of standardized test data. In short, we believe that the evils attributed to standardized tests by opponents of these programs are illusory.

Standardized tests are not only benign, they are vital national educational resources. I have argued this in the historical comments above. Let me now give an example of current interest. I has been evident to college and university teachers (among others) ever since the turbulent 60's that something profound and disturbing was happening to the academic qualifications of incoming freshmen. Classroom teachers have been relating anecdotal evidence of a decline in the apparent ability of entering students for over a decade. But the evidence was local, or at least locally observed. The first reaction of an individual teacher or an individual department to the perceived decline in student ability was to suspect that university standards must be creeping up. Teachers wondered whether they were gradually expecting too much of their students; deans, alarmed at depressing grade statistics and increased failure rates, pressured academic departments to mind their grading standards; allowances were made to restore the familiar bell curve to its normal position. There was no national consensus to ease grading standards. There were only thousands of individual adjustments made conciously or sub-conciously to meet trends which were felt but only dimly perceived. The net effect was as smooth a national program of grade inflation as could have been orchestrated by the most efficient and influential Educational Commission.

Then it was noticed that SAT scores had been declining for quite a number of years. This observation first announced the possibility that, for whatever complex reasons, the level of training and ability of entering college students was in a real decline. A flurry of statistical research on trends over time and across populations (male/female, black/white, urban/rural. . .) and a cacophony of discussion and debate exploded on our nation. Some insights on trends in our educational system and in our society have emerged. Among other things, we suddenly recognized rampant grade inflation for what it was. Standards have begun to be tightened. Some basic changes in both secondary and post secondary education are being made to reverse a trend that, once recognized, was frightening.

The point of this anecdote is that it was precisely the testing programs that had been in place for more than twenty years, involving tests that were carefully equated over time and population, that enabled us to see clearly a trend that was invisible in our day-to-day academic experience. The tests did not immediately reveal the nature of the trend, nor did they provide solutions to the problem. But they were a vital data source - an important national resource - that served our nation well in a crisis.

What does this have to say about the Weiss Bill? The answer is clear when one realizes that one result of H.R. 4949, possibly not foreseen by its authors and supporters, would make it virtually impossible to standardize tests. Such standardization, called "equating," involves comparing different populations of test-takers by studying how they perform on selected test items which must be repeated in the administration of the tests to these populations. If the test populations being compared exist at different times then this involves repeating test questions at different times, often many times over a period of years. This would clearly be impossible if all test questions were made public after each test administration.

Without standardization of tests students would again be prisoners of class, location, and time. Colleges and universities would once again be forced to rely on established "old-boy" networks giving preference to preparatory schools whose programs are well known to the individual admissions officers and using other tried and true but restrictive screening devices. American education could be set back thirty years by the removal of standards from standardized tests.

As we all know, the State of New York has passed a "Truth in Testing" law which requires disclosure of standardized tests. It is legitimate to ask why, in the light of the arguments presented above, the SAT's, the Graduate Record Exams, and other standardized tests will be continued in New York. The answer lies in the fact that the New York law specifically exempts certain types of tests (e.g. the CEEB achievement tests) and certain parts of other tests (the equating sections of SAT's) from the disclosure requirement. Without these exemptions, all of these tests would be frozen out of New York State entirely and New York students would be placed at a distinct disadvantage, vis a vis their counterparts in the rest of the nation, in admission to colleges, universities, and professional schools.

In fact, several testing organizations, specifically those offering tests used for admission to law and medical schools, have announced that their tests will henceforth not be given in New York. These tests, like many others, are of the type for which equating is not possible by means of special "non-operational" sections specifically exempted by the New York law.

This serves to illustrate the havoc that can be caused by a hastily constructed disclosure law, no matter how good the intentions of the framers.

In its present form the Weiss bill contains a disclosure requirement patterned after, but more inclusive, than the requirement in the New York law. It would therefore be a far more serious blow to standardized testing. But the fact that SAT's and GRE's could be saved in New York only by the specific exclusion of these particular tests is adequate testimony that the operative disclosure provisions are poorly conceived. Already we are learning of other tests that should have been excluded. How are the citizens of New York to know what other testing programs, present or future, will be crippled or made impossible by the indiscriminate requirement of disclosure. Expanded to a national level such a blow to a major educational resource would be catastrophic.

In testimony to your committee, others have described additional ways in which the Weiss Bill would harm education in the United States. Others have described the strong efforts of testing organizations to disclose as much information about tests and testing procedures as possible within the constraints of standardization. Still others have described the strong efforts of the educational community, through such organizations as the American Council on Education, to review and evaluate all standardized testing programs in the light of present-day concerns (which we all agree are quite valid in our free democratic society). We wish only to add our strong plea that the disclosure clause in the Weiss Bill not be allowed to become law. This is why we oppose the Weiss Bill. If it is defeated then we can perhaps consider other legislation incorporating some of the beneficial aspects of H.R. 4949 and H.R. 3564.

Thank you for your careful consideration of our concerns and for all of your efforts to improve educational opportunities for the youth of our nation.

PREPARED STATEMENT OF JACK J. STOCKTON, D.V.M., PRESIDENT, ASSOCIATION OF AMERICAN VETERINARY MEDICAL COLLEGES

"Standardized" examinations are given for a number of reasons. A few of the very important ones are: (1) to establish "norms" and identify those who, for whatever reason, are at the extremes; (2) to provide some long-term and continuing basis for comparisons which may be useful in evaluating the total system; (3) to identify "areas" of strength and weakness in the examinees both individually and collectively; and (4) to provide another element in the process used for selection. Admissions groups need all the help they can get in making sound judgments on selections to highly competitive programs, and licensing boards must determine competency to perform and render reliable services to the consuming public. In a purely academic setting, the educational value of providing feedback on examinations has proponents and opponents, but in that setting there is general recognition that such procedures do have educational value. However "Standardized Testing" as currently conducted, whether it be for entry to the university or for professional licensure decisions, has different objectives. The state and national dimensions of these examinations also imply an entirely different set of circumstances than that for an examination in a given course in a given educational institution, and detailed feedback is not essential.

Specifically, the Association of American Veterinary Medical Colleges speaks against the proposed legislation because:

(1) We are not convinced that such proposals would solve any ills that are perceived as currently existing. In fact, such legislation would be complicating and counterproductive to the overall objectives of standardized testing as currently practiced.

(2) The body of knowledge is extensive and sampling is essential. As major and important concepts are tested for in examinations, new test items would tend toward the trivial if questions and answers were routinely published.

(3) Flexibility would be greatly reduced. Students now have an opportunity to take a number of examinations in varied locations at various times. For example, the Veterinary Aptitude Test (VAT) is currently offered 5 times per year. There would be a move to many fewer offerings per year and in the instance of VAT probably to once per year.

(4) The impact would be inversely proportional to the size of the group affected. For example, most high school students in the United States take the SAT and/or ACT Test, and there is a relatively low development cost per student as compared with those tests for many, many fewer examinees, e.g., human medicine and MCAT and veterinary medicine and VAT where the development cost would increase significantly.

(5) There would be a great waste of professional time in the redundancy of new examination construction, detracting from the ability to evaluate, analyze, and improve.

(6) The tendency to produce a "levelling effect" in test performance is a grave concern, since the advantages of innate ability and concerted study effort on a broad front would tend to be negated.

Fairness, objectivity, and capriciousness are not in contention as these are fundamental and obvious concerns of society. Examinees need to know the nature and use of examinations and to be assured that test scores accurately reflect performance. Feedback on areas of strength and weakness should be available. We believe testing services are sensitive to these matters and are not just the money-mongers some try to make them out to be. However, the profit motive is as real in providing services as it is in selling hard merchandize, and the free enterprise system, if it is to obtain, cannot be battered and beat by the very hand that it feeds.

Evaluation is as fundamental to the educational system as the jury is to the judicial system. An examination, which is an evaluation device, is an exercise to determine a person's knowledge, ability, and qualifications in a prescribed area which may be rather broad or rather limited in scope. If a person has both the questions or problems and the answers, there is in effect no examination, and the fundamental objective of the whole process--evaluation--is negated. That the system is imperfect is recognized and admitted, but so is the political system, the judicial system, or any other system devised by imperfect people. Before the system is radically manipulated it should be understood as it is being used and with the full recognition of what radical changes will accomplish.

As participants and educators in a health profession field, we support the concept of standardized testing. Concurrently it should be emphasized that such tests are not the only criterion used in arriving at judgment decisions which must be made. Rather, test results are combined with other academic and sociological data to provide a breadth of input in helping to evaluate ability and motivation. Discriminating elements need to be honed and refined and not blunted or made more diffuse.

The Association of American Veterinary Medical Colleges can see no benefit to students, the general public, educational institutions, or the educational testing"industry" from these bills, and we question whether the bills would have the effects claimed by their proponents. We believe it is necessary to reject these bills in order to preserve the systems of graduate medical admissions and professional licensing as they now exist.

Respectfully submitted,

Association of American Veterinary
Medical Colleges

Jack J Stockton

Jack J. Stockton, D.V.M., President

Prepared Statement of the National Council of Architectural Registration Boards

1. *The National Council of Architectural Registration Boards* - The National Council of Architectural Registration Boards (NCARB) is a non-profit corporation whose only members are the architectural registration boards of the fifty states, the District of Columbia, Puerto Rico, Guam and the Virgin Islands. NCARB's principal functions are to foster uniformity of laws relating to the regulation of architectural registration and practice in order to facilitate the free flow of professional practice in interstate commerce. Its activities are more fully described in the Appendix attached and entitled, "How Architectural Registration Boards Work Together."

2. *The Architectural Licensing Examinations* - The States and Territories whose architectural registration boards comprise NCARB have, beginning in 1897, all enacted statutes regulating the practice of architecture and creating the state registration boards to enforce the respective state laws. Common to the registration procedures of all of the States and Territories is the requirement that all candidates for architectural registration pass certain licensing examinations which are now prepared by NCARB with state board participation and are adopted for use by all state boards, a process which has been judicially recognized and approved. *Henkes* v. *Fisher*, 314 F. Supp. 101 (1970), *aff'd.*, 400 US 985 (1971).

The Professional Examination consists of two parts. Part A is a design examination requiring candidates to demonstrate their competence to produce a graphic solution to an architectural design problem. Part B is a two-day, 16-hour multiple choice examination which poses to the candidate actual architectural practice situations in which his or her abilities to exercise value judgments are tested and evaluated. This examination is given in four sessions covering the following fields in the context of a specific building project:

(a) *environmental analysis* - covering the natural and manmade surroundings, relevant laws, and the community setting;

(b) *architectural programming* - covering overall planning of the project;

(c) *design and technology* - covering detailed design of the project; and

(d) *construction* - covering contract administration, construction management, construction methods, and materials.

Candidates must achieve a minimum score on each of the four parts in order to pass.

Candidates without an accredited degree are also required to take the Qualifying Test, a two-part 22-hour examination divided into the following sections: architectural history; structural technology; materials and methods of construction; environmental control systems; and principles of site planning and architectural design. All sections are multiple choice. As with the Professional Examination, a minimum score for each separate part must be attained.

Through NCARB, the content of these exams is reviewed continuously by representatives of boards around the country. Educational Testing Services of Princeton, New Jersey, serves as testing consultants. In 1977 an independent panel of internationally recognized architects in practice and in education reviewed the objectives of the Professional Examination in order to verify that the examination was reasonably related to the areas of knowledge, skill and ability being tested. This process will be repeated periodically to ensure that the examinations do their job.

Part B of the Professional Exam is written by the members of NCARB's Professional Examination Committee. For the most part, the exam writers are practicing architects; a few of them are educators. Nearly all are drawn from the NCARB's member registration boards, with geographic representation through the country.

About 30 architects serve as exam writers each year. They work without compensation under the supervision of a Chairman and four Coordinators. A coordinator directs the work of seven or eight exam writers on each of the four parts which comprise the Professional Exam. The NCARB Examinations are more fully described in the attached Appendix entitled, "The NCARB Examinations--A Summary Report of Their Preparation and Setting of Passing Scores."

3. H.R.3564 - NCARB urges that H.R.3564 be given unfavorable action by the Congress. H.R.3564, a bill introduced by Congressman Gibbons of Florida, "To Require All Educational Admissions Testing Conducted Through Interstate Commerce, and All Occupational Admissions Testing (Which Affects Commerce) to be Conducted With Sufficient Notice of Test Subject Matter and Test Results, and for Other Purposes," is flawed in three basic respects: first, it establishes federal jurisdiction over the entire range of all occupational licensing which heretofore has been a local matter left to the individual states; second, it establishes federal occupational licensing policies (and the precedent for such policies) and ousts the states from enforcing inconsistent or contradictory occupational licensing policies; and third, as with nearly all sweeping declarations of broad principles, this particular bill establishes requirements which are ambiguous and difficult to properly administer and which invite needless and disruptive litigation on the part of disappointed candidates for architectural licenses -- of whom there will always be some.

A. Creating a federal jurisdiction over occupational licensing is unwise.

Since the early days of the Republic, it has been a recognized and accepted function of state governments to regulate activities which, even though formally private in nature, nonetheless affect the public health, safety or welfare. One aspect of this role has been the regulation of the professions, whose members are properly considered to have special responsibilities to the public as well as to the individuals receiving services. The essential rationale and standard for such regulation was set forth by the U.S. Supreme Court in *Dent* v. *West Virginia*, 129 US 114, 122 (1889), when the Court said:

> The power of the State to provide for the general welfare of its people authorizes it to prescribe all such regulations as, in its judgment, will secure or tend to secure them against the consequences of ignorance and incapacity as well as of deception and fraud. As one means to this end it has been the practice of different States, from time immemorial, to exact in many pursuits a certain degree of skill and learning upon which the community may confidently rely, their possession being generally ascertained upon an examination of parties by competent persons, or inferred from a certificate to them in the form of a diploma, or license from an institution established for instruction on the subjects, scientific and otherwise, with which such pursuits have to deal. The nature and extent of the qualifications required must depend primarily upon the judgment of the State as to their necessity.

NCARB does not know all of the occupations for which all of the states now administer some form of "aptitude or achievement examination, whether written or oral, and [including] any objective multiple choice, machine scored, essay, practical, performance, or demonstration examinations"--the definition of an occupational admissions "test" to which H.R.3564 applies. We suspect, however, that the list is nearly endless, and growing each year. The precedent established by enactment of this bill would therefor extend federal jurisdiction in a massive reach over a multitude of professions, occupations and trades.

We do not hold any particular brief for this growing list of regulated occupations, and believe in each instance (before subjecting an occupation to regulation) that the states should answer the twin questions of how such regulation will protect the consumers of services rendered by the practitioners, and how such regulation will protect the public at large. We believe that in the case of architectural practice, appropriate regulation both protects consumers who might otherwise be victimized and the public who might otherwise be injured.

But the particular standards, including testing standards, which can most effectively protect the public interest is a matter far better left with the states. Establishing standards for entry into professions and occupations is a subject particularly appropriate for exercise of the states' traditional police powers. The diversity of different regions' needs and the differing perceptions of the public interest can better be accommodated at the state level. On the other hand, establishing uniform federal standards regulating occupational entry seems almost certain to require that there be a single view of the public interest that submerges the different regional needs and the perceptions of people in a given locality.

B. The new federal jurisdiction would supplant state jurisdiction over subject matter now regulated.

H.R.3564 addresses two subject matter areas, one -- education testing -- not now generally subject to any governmental regulation, and the other -- occupational testing -- presently subject to extensive governmental regulations. For all of the reasons mentioned elsewhere, we question the wisdom of this bill, but if the Congress feel that there is sufficient need to broaden federal jurisdiction testing generally, we suggest that it give primary attention to such testing as is now subject to no regulation.

Architectural licensing tests are clearly covered by this bill; and they are now the subject of state law in all the states. This bill is not merely bad precedent; at least in two respects it could supplant inconsistent state law policies. In Sec. 6 (a)(3), the bill requires that each applicant be provided with a written notice containing "the margin of error or the extent of reliability of the test, determined on the basis of experimental uses." Passing the ambiguity of this mandatory requirement, it appears at the least to require that all occupational tests be verified in some manner in terms of their ability to predict or measure knowledge or aptitude. And being mandatory, if an architectural registration board had not verified this by experimental usage -- and it is unclear how experimental usage of different yearly tests could so verify -- then it could not give notice of this required information, and could therefor not give the examination. This requirement presents a particularly acute problem regarding the design portion (Part A) of the Professional Examination for architects. We believe architects must be minimally competent to express themselves graphically, and testing for such design competence is quite clearly "practical, performance, or demonstration examination" covered by the bill's definition of "test" (Sec. 2(3)). Yet we are at a loss to see how state registration boards can determine "margin of error or extent of reliability on the basis of experimental use," and give the required notice to applicants for architectural licensing.

In a more direct manner, H.R.3564 prohibits a particular method of grading tests. (Sec. 6(c)) In this respect also, the bill would supplant all state policies which are inconsistent.

C. The bill's ambiguities invite needless and disruptive litigation in the federal courts.

The most critical ambiguity of H.R.3564 is found in Sec. 6(a)(3) noted above. None of these terms reflect a common body of meaning and general understanding in the testing field generally; they have even less meaning when they are sought to be applied to particular occupational tests such as architectural tests. Moreover, there is no federal authority given rule-making responsibilities to enlighten the state occupational licensing boards through regulations as to the proper meaning of the "reliability" requirement. Instead, individual Federal

District Court judges will be left to decide the Congressional intent on a case by case basis. And in almost all instances, the decision will have to be made by the judge (and responded to by the state licensing board) under the severe time pressures of an application for permanent or temporary injunction. (Sec. 7(a))

We cannot state that no grading errors have ever occurred in grading the architectural examination, or that the test format or questions cannot be improved. Each year literally hundreds of hours are given without compensation by practitioners, educators and others who are the state officials serving on NCARB's member state boards to assure that examinations accurately and adequately test candidates' knowledge, skill and ability in applying professional level judgment to architectural problems.

There is a substantial body of case law which over the years has developed to protect candidates for occupational licensing against arbitrary and capricious actions or other violations of their due process rights. The majority view has been that in the absence of arbitrary or capricious acts, failing candidates' rights are sufficiently protected if they may retake the failed examination. *Tyler* v. *Vickery*, 517 F. 2d 1089, 1101 (5th Cir. 1975) *cert. den.* 426 *US* 940 (*1976*); *Singleton* v. *La. State Bar Ass'n.*, 413 F. Supp. 1092 (E.D. La. 1976); *In Re Mead*, 1977 Mass. Adv. Sh. 599 (Ma. 1977). *See*, *Detroit Edison Co.*, v. *N.L.R.B.*, 47 U.S.L.W. 4233 (1979); *Henkes* v. *Fisher*, *supra*.; *Whitfield* v. *Illinois Bd. of Law Examiners*, 504 F. 2d 474 (7th Cir. 1974); *Chaney* v. *State Bar of California*, 386 F. 2d 962 (9th Cir. 1967). NCARB submits that this evolving body of law offers more appropriate protection for disappointed candidates then does H.R.3564.

HOW ARCHITECTURAL REGISTRATION BOARDS WORK TOGETHER

INTRODUCTION

This document has been prepared by the architectural registration boards of the United States in an effort to acquaint state legislators and members of state executive departments with the process by which architectural registration boards work together. It explains the inter-relation of the state boards and the National Council of Architectural Registration Boards ("NCARB") which serves as a national forum for its member boards.

THE NATURE OF ARCHITECTURAL PRACTICE AND ITS EFFECT ON REGISTRATION

It is important to appreciate, at the outset, the way in which architectural practice in the United States differs from that of most of the other learned professions and how that difference has affected registration.

While the regulation of the practice of architecture, like other learned professions, is remitted to the states by tradition going back to the nineteenth century, the architect is typically involved in interstate commerce in a way unparalleled by other professional groups. It is true that lawyers, doctors, and accountants require some measure of reciprocal licensing in order to permit the lawyer, doctor or accountant to move his practice from one state to another (without the necessity of submitting himself again to the full, initial licensing process). These professions rarely, however, require the capacity to move routinely across state lines. Architects, on the other hand, customarily do their work in their base state and surrounding states; and America's larger architectural firms have a practice which is national, if not international, in scope.

Of course, each particular state is concerned with the quality and safety of buildings constructed within its borders. Each state has as keen an interest in examining the qualifications of an architect from another state coming in to do a project as it has with regard to the qualifications of an architect resident within the state.

The process by which an applicant is admitted to registration in his base state is appropriately substantial and exacting. To require that every architect seeking to design a project to be built in a state other than his base state comply with that substantial procedure would seriously obstruct the capacity of architects to move from state to state. Both architectural registration boards and the profession at large early recognized that there was no place in the system for the local protective instinct which sought to obstruct the free flow of practice from state to state. In general, there has been nearly unanimous recognition of the notion that it is in the interest of the public as well as the profession to facilitate interstate practice.

Thus, each state was charged by its legislature to protect the public health, safety, and welfare by examining the qualifications of persons who were to be permitted to design structures within that jurisdiction. At the same time, there was a clear understanding that the public interest required that the practicing architect be free to move from state to state in order to practice his profession. The first solution to this dilemna was to provide in state statutes that an architect registered in his base state could come into another state if the base state permitted the second state's architects similarly to have access to practice in the base state and if the standards for registration of the base state were no less than the requirements of the second state. Such a solution broke down in two respects. First, it was exceedingly difficult for an individual state to scrutinize and evaluate the requirements of every other state. Second, it was difficult to assess whether another state was opening its doors to architects from the host state. A state board heard of aggravated cases only, and often those cases overshadowed the good-faith reciprocity which may have existed.

A much more effective solution was to provide a national clearing house, controlled by the boards of the several states, which would evaluate and recommend candidates seeking registration in a second state. This solution, however, posed still another problem. If there was an infinite variety of standards among the fifty-five jurisdictions, could any state board, charged with protecting its citizens, allow a national clearing house to make judgments concerning the equivalency of standards?

The registration boards ultimately succeeded in resolving this problem to a large extent by creating the NCARB as a forum for the joint development of normative standards for the nation. Thus, the member boards as a group, acting through the NCARB, study and recommend standards for training and education which, when adopted by the member boards individually, become the NCARB standards for granting its Certificate—a document evidencing the NCARB's affirmative recommendation to a member board that the Certificate-holder is qualified. In fact, the process has worked. All states use the national examinations developed by NCARB. All states recognize the NCARB Certificate as evidence of an applicant's qualification to practice. Many states have adopted the NCARB

6/77

training and educational standards as their own standards for registration.

WHAT IS NCARB?

In 1919, the architectural boards of a handful of states, recognizing the problem posed by inter-state practice, met to discuss the formation of a national organization. These discussions resulted in the establishment of the National Council of Architectural Registration Boards in 1920.

NCARB, located in Washington, D.C., is a non-profit, membership organization established under the laws of the State of Iowa, comprising the registration boards of the fifty states, the District of Columbia, the Canal Zone, Guam, Puerto Rico, and the Virgin Islands.

It is important to emphasize that NCARB's only members are the legally constituted architectural registration boards. Only the boards formulate the rules and policies of the NCARB, and elect its directors and officers. To be elected a director or officer, a candidate must be currently serving on a member board, although he may subsequently be reelected or elevated to a higher office at a time when he may no longer be serving on his state board.

The chief activities of the NCARB can, for convenience, be divided into two general areas. The first area comprises those services rendered by the NCARB to its member boards, including the development of the registration exams to be administered by member boards, the development of recommended education and experience standards which may be adopted by member boards as the standards to be required of candidates for registration, and the development of recommended standards of professional conduct which may be adopted by member boards pursuant to the rule-making power generally found in registration statutes.

The second category comprises services to registered architects, the chief of which is the maintenance of a nationwide system, based upon the Certificate issued by the NCARB, by which an architect registered in one state is admitted to practice in another state. This system, perhaps unique to the profession of architecture, has evolved over the years and represents the principal path by which architects are able to conduct their practice beyond the borders of their base state.

In its most recent fiscal year, the dues collected from its member boards comprised less than two percent of NCARB's gross revenues. NCARB's charge for the national examinations and associated charges comprised approximately thirty-eight percent of the Council's gross revenues. Nearly sixty percent of the gross revenues of the NCARB came from registered architects using the Council's services in connection with interstate practice. These figures underline the rather unique position of the Council when compared with similar national councils of governmental organizations: while the entire control of the NCARB is in the hands of its member state boards, over half of its revenue is generated from the services it affords private practitioners.

HOW NATIONAL STANDARDS ARE DEVELOPED

NCARB engages in extensive work by committees to develop the recommended standards which were discussed above. The committees comprise members of NCARB's member boards. The Committee on Procedures and Documents re-examine each year the standards for training and education; the Committee on the Intern Architect Development Program is a major effort to re-examine the process by which young applicants for registration develop adequate training experience; the various examination committees deal with the Qualifying Test and the Professional Exam; and the Committee on Professional Conduct is presently making recommendations to all member boards concerning a Code of Professional Conduct which the member boards may adopt by regulation. In this fiscal year, thirty-seven states have members of their boards participating in this committee work. The States without participating committee members this year will doubtless have participants in the next few years, for it has been the conscious practice of the NCARB Board of Directors, who represent the member boards and who come from each geographic area of the United States, to constitute important committees in such a way as to be sure they reflect a broad geographic cross-section of member boards.

During the year, the committees deliberate on their areas of responsibility. In cases where outside consultants are necessary, (*e.g.*, testing experts, legal counsel, educational experts, and the like) these consultants are retained by NCARB to assist the committees. The work of the committees is reported periodically to the NCARB Board of Directors. Every important policy consideration, including any important change in the standards to be recommended to the member boards, passes under the scrutiny of the Board of Directors.

In addition, policy matters are reviewed in the separate meetings of the Regional Conferences of the NCARB. Each member board is also a member of a Regional Conference, which is a geographical grouping of boards; there are six Regional Conferences. The Regional Conferences hold at least two meetings yearly, which are crucial to the process. These regional meetings are frequently a source of new policy proposals as well as an occasion for study of outstanding matters. Typically, national officers and staff personnel attend each regional meeting. Their task is to provide a detailed picture of proposals which are expected to be submitted for vote at the Annual Meeting. These proposals are discussed and frequently modified at the regional meetings.

At least thirty days prior to the Annual Meeting in each year all matters to be voted on must be sent to the member boards. This requirement enables the member boards to deliberate themselves on the questions to be raised at the Annual Meeting.

At the Annual Meeting new policies to be voted on are presented in detail. The chairman of the committee responsible for such policies is given ample time to present the proposal to the entire Annual Meeting. They are informally discussed in regional meetings held at the time of the Annual Meeting. Frequently the committee chairmen or one of the national officers is invited to join the regional meeting to answer questions raised by region members. Thereafter, the policies are debated in plenary session.

When the issue comes to a vote, each member board has a single vote. It has been the tradition of NCARB that no major policy change is put into effect when supported only by a narrow majority. All such changes have been held in abeyance until there is substantial consensus in favor of the change. The membership is always sensitive to the fact that a change in the standards for the NCARB Certificate effectively changes the rules by which a registered architect is free to practice in another state; if a single state refuses, as a result of the change, to accept the NCARB Certificate, the entire structure of interstate registration may be undermined.

SUMMARY

The foregoing underlines the important role of NCARB as the forum through which its member boards affect national registration policy. At a time when responsible state budgetary authorities are seeking ways to reduce the cost of government, it is useful to emphasize the unique utility of the NCARB for its member boards. As an example, it is commonplace for items of "out-of-state-travel" to be the first victim of budget reviewers. While many such inter-governmental agency meetings are useful only as an educational experience (exchange of ideas, staying abreast of current developments), the NCARB meetings inevitably affect the policy of registration in each member boards' sovereign state. No state can afford to be absent when changes in standards are discussed and adopted. While it is true that a particular state may refuse to adopt the new standard when considering an application from another state, if it does so persistently, it runs the strong risk that its own registrants will be barred from practice in other states. In short, the delicate national system of mutual respect and comity among the registration boards of the sovereign states survives only because, to date, substantially all of the states have recognized the importance of the decisions taken by member boards at NCARB meetings.

As another example, many state government officials question their board as to why it should defer, in any way, to the judgments of a substantial number of the boards of sister states. Each state is sovereign and can make its own judgment as to how best to protect the health, safety, and welfare of its habitants. No one denies the validity of that proposition; but no one can deny either that the free flow of professional practice throughout the country requires a sense of mutal respect and comity among registration boards. It is through the forum of NCARB that the delicate balance is maintained.

APPENDIX D – THE NCARB EXAMINATIONS

A SUMMARY REPORT ON THEIR PREPARATION AND SETTING OF PASSING SCORES

1. BACKGROUND

Before the uniform multiple-choice testing format was developed, Registration Boards wrote their own individual examinations. These were generally of the essay type. Each Board graded its own papers and established its own criteria and standards to determine passing papers. The subjects covered in the examination followed a syllabus recommended by the NCARB, but the content of each examination was determined by each Board. The passing scores for each section of the examination varied from 60 to 75; however, passing was determined on the basis of the total examination and of achieving a percentage (e.g. 60, 65, 70 or 75) of a possible 1,000 points.

In the late '50s, NCARB began investigating the use of the multiple-choice format for the seven-part, four-day, 36-hour examination. The first multiple-choice tests were used in 1962, and by 1965 all portions of the seven-part test except for the two graphics tests were cast in the multiple-choice format.

The required passing scores for these tests were determined by the members of the Member Boards meeting at the 1965 NCARB Annual Meeting in Washington, D.C. The procedure called for the delegates present to evaluate the questions in each of the tests and to rate the difficulty of each question for the minimally qualified candidate. In addition, each delegate was asked to indicate his professional opinion as to the percent of the total number of questions the candidate should answer correctly in each test in order to pass. All data were collected and evaluated, and the exam passing standards were established.

In 1968, further research by NCARB led to the development and implementation of what are now known as the Qualifying Test and the Professional Examination. These two examinations became effective in 1973. (Note: The Qualifying Test was initially entitled the Equivalency Examination.)

2. EXAMINATIONS

The Qualifying Test is designed to test the candidate's knowledge, skill and ability in the fundamentals of the art and science of the practice of architecture.

The Professional Examination is designed to test the candidate's knowledge, skill and ability in applying professional level judgment to architectural problems and their ability to synthesize all technical, programmatic and aesthetic aspects in the resolution of an architectural design problem as related to the practice of architecture and the protection of the public.

A. Test Development Background

1. *Qualifying Test:*

Before any subject areas were defined or any test questions written, an exhaustive study was made to determine which subjects form the core of knowledge, skill and ability necessary for the practice of architecture. The Committee assigned to this task concluded that the test should be designed to elicit a level of knowledge, skill and abilities equivalent to that of a student about to be graduated from an accredited program in architecture.

Preliminary content specifications for the four subject areas of the Qualifying Test (History, Structures, Materials and Methods, and Mechanical) were prepared by the Committee. These test specifications were then distributed to all schools of architecture with accredited programs for their review and comment as related to the content and depth of coverage of each subject as taught in the schools.

Additionally, faculty members from the schools were requested to submit copies of their final examinations as well as a number of multiple-choice test questions for the Committee's review and possible use. The information collected from the schools were analyzed and the final test specifications were prepared. The writing of test questions was then undertaken complying with these specifications.

The NCARB Committee writing the Qualifying Test consists of architects, educators and Board members. Two-thirds of this Committee is made up of architect-educators recognized in academia for their abilities and knowledge of their specialties. The architect-Board members are noted for their expertise in particular subject areas and bring to the Committee the concerns of registration Boards as a policy-setting factor. The Committee has sixteen active members who meet twice a year in formal sessions and also contribute considerable personal time in writing test questions at home.

2. *The Professional Examination*

The Professional Examination, multiple-choice sections, (identified as Section B) progressed to its present format after several years of intensive study and development. The subject areas of the Examination were derived after a study was made of the practices, procedures and services offered by architects, and an analysis and projection of what architectural practice and public needs might be in the future. The schools of architecture were consulted for their appraisal of the future of architecture in order to develop realistic and appropriate subject areas and test specifications.

A mock Professional Examination was developed and the Member Boards reviewed the concepts, subject areas and sample test questions. The procedures, format, subject areas and schedules for the new examination were adopted in 1972 and the first official test was given in December 1973.

The initial NCARB Committee of forty persons charged with the development of the Examination was comprised of architects, educators, generalists and interns. Over the years, this Committee has been effective with fewer members and now numbers twenty-eight persons comprised of architect-Board members and architect-educators. This Committee meets three times during the year and each member spends a great deal of personal time in writing questions, reviewing TIP material and *taking the examination.*

3. EDUCATIONAL TESTING SERVICES (ETS) INVOLVEMENT IN NCARB EXAMINATIONS

Educational Testing Service of Princeton, New Jersey, and Berkeley, California, has assisted NCARB in the preparation of examination materials for more than twenty-five years. Their testing specialists instruct the NCARB Committee members on how to write multiple-choice questions; edit the questions for clarity, punctuation, grammar, etc.; score the exams; prepare rosters of the candidates' performance; and accomplish the statistical and analytical work related to the whole examination as well as to individual questions.

ETS neither attempts to write examination questions nor to establish the required minimum scores for passing; however, they do provide background data and guidance on accepted testing procedures and provide the technical expertise and administrative workforce for the NCARB prepared testing instruments.

4. EXAMINATION WRITING

Questions in every NCARB examination are written by architects or educators who teach architectural subjects. Questions are reviewed, criticized, rewritten and either accepted for use or rejected. The questions are written to cover the subject areas as specified. Each question in all NCARB tests except Structural Technology are valued at one point each. In the Structural Technology test, fifteen questions are valued at five points each because they are calculation questions, taking more time to answer; therefore, these questions have a higher point value.

Within every test a series of questions appear that have been used in previous tests. These questions are used without modification for the purpose of comparing the difficulty of the current test with previous tests and to compare the ability of the total candidate group with the group taking previous tests. These questions are called "equaters" in that they provide statistical data that aid in establishing the required minimum raw scores for passing the test thereby enabling the maintenance of uniform passing standards for each test.

SETTING OF MINIMUM REQUIRED SCORES

Of critical importance is the establishment of the initial requirements for the demonstration of minimum competence in a given subject area. This is the most difficult and agonizing aspect of the entire examination process. After the test has been given and a significant number of answer sheets have been scored, those few questions (items) in the test that have poor performance statistics are analyzed by the Chairman of that portion of the test. Then a decision is made on whether to retain the item, to reject it or, in rare cases, to change the answer (key).

The next step is to analyze the test with regard to its difficulty as compared to previous versions of the test. A test's difficulty is determined by tabulating the difficulty values of the test questions and dividing that total by the number of questions in the test. Difficulty is determined by observing the performance of candidates and translating that performance into numbers through use of formulas standard to the testing industry. A value of 11.7 on a scale of 0–20 represents middle difficulty. NCARB tests strive to be in the range of 10.0–13.0 in difficulty rating. Also, the candidates, as a group, are evaluated for their overall ability and then compared to the previous candidate group ability. Candidate group ability is assessed by measuring performance of the brightest candidates of one group to the ability of the brightest candidates from previous groups. These analyses are accomplished through the incorporation of the "equater" questions (see Item 4 above).

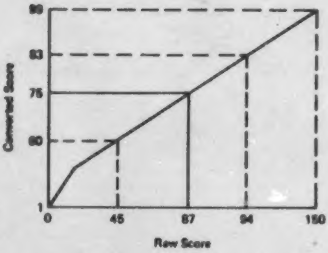
Regardless of the number of questions in a test or the length of time allowed to answer the questions, the passing score of a test can be set that is equivalent to that of previous tests through application of the test difficulty and group ability data. Through this process, once the initial levels of performance have been set, each new test applies uniform standards in relation to its predecessors with regard to the candidate ability required to attain the minimum passing score. Thus, it becomes possible to require that candidates perform at the same standards as previous candidates.

No test is considered more important than any other; all are considered equal in value. One portion of the Qualifying Test—Section B—has a number of questions which carry a higher point value. This is sometimes called "weighting". The questions in the Structural portion of the test that are "weighted" are the calculation-type questions. The purpose for "weighting" such questions is to ensure that the candidate will attempt to answer them and, also, to reflect the importance of answering such questions in order to pass.

Members of the Examination Committee must analyze the data generated from the test administration as noted above. If the test under consideration is a continuation of previous versions, "equating" data are utilized in the setting of the required minimum raw score (number of questions answered correctly). If the test is the first of a new form, historical data from previous tests on similar subjects, or "pre-test" data, will be considered and incorporated into the procedure.

After careful study of the data, the Committee makes a determination on behalf of public health, safety and welfare as to the level of performance required to demonstrate minimum competence in the subject area of the test. The minimum required raw score established in this manner reflects the Committee's professional judgment of the expectations of the public and the profession with regard to minimum competence. Once the minimum required raw score is established, a conversion table is developed through which raw scores can be converted to a scale ranging from 1 through 99. In the conversion table, the required minimum score is set to equal, or convert to, 75. The number "75" is simply an arbitrary number and does *not* represent a percentage of the total number of questions in the test. The number "75" is the converted score required to pass the test.

(On a test where raw scores can range from 0 to 150, and a raw score of 87 is determined to be the minimum required score to pass, 87 is set to convert to 75. A raw score of 94 might convert to 83; and a raw score of 45 might convert to 60.)



The difficulty of any test has a direct relationship to the required minimum passing score. The tests are not intended to rank candidates (that is, to find out who is best, average or worst). The tests are designed to determine if a candidate has demonstrated minimum levels of knowledge, understanding, skill and ability in the subject area. Since each test covers a number of salient points about the subject of the test that are related to principles, theories and their application, there may be several test questions on single aspects of the subject. Consequently, several opportunities are given to the candidate to respond to the principles and theories of the test subject. The test is not designed to require expert status on any subject; its purpose is to allow the candidate to demonstrate minimum competence.

Samuel T. Balen
Director of Professional Development

nabp
National Association of Boards of Pharmacy
1 East Wacker Drive • Suite 2210 • Chicago, IL 60601 • 312/467-6220

July 20, 1979

Chairman Carl Perkins
Subcommittee on Elementary, Secondary,
and Vocational Education
Committee on Education and Labor
B-346-C Rayburn House Office Building
U.S. House of Representatives
Washington, D.C. 20515

Dear Mr. Chairman:

In response to the needs of our member boards, the 57 state boards of pharmacy in the United States and Canada, the National Association of Boards of Pharmacy (NABP) is now the producer of a standardized and uniform entry-level licensure examination used in 49 American jurisdictions. Our examination, entitled NABPLEX (for the "National Association of Boards of Pharmacy Licensure Examination"), would fall under the scope of "The Truth in Testing Act of 1979," H.R. 3564, as introduced by the Honorable Sam M. Gibbons, and referred to your subcommittee.

In the sponsor's "Dear Colleague" letter of May 3, Mr. Gibbons said, "I would appreciate your suggestion on how this bill can be improved with its overall purpose in mind without unduly burdening testing firms, professional and occupational licensing boards or educational institutions." The proposed Act is basically a good and useful instrument, yet some provisions would be injurious to the methods used in constructing and administering a secure and non-discriminatory examination. Specifically, Section 6(b)(4) is too broad and would permit examination candidates to request scores of other candidates, copies of test items or other test information which would contradict good testing procedures and good public policy.

In addition, Section 6(c) appears to be aimed at eliminating norm-referenced score interpretations. However, the language eliminates the use of statistical treatment and reporting of scores based on the relative distribution of scores. It would also eliminate the use of average scores which are based on scores of other test subjects which have been statistically treated on the basis of score distributions. This language is awkward and would militate against the use of

Honorary President Vincent Tassone, Ogden, Utah
President Earl W. Chase, Washburn, North Dakota
President-Elect Joseph G. Shalinsky, Kansas City, Kansas
Vice President J.H. Voige, Frankfort, Kentucky
Treasurer Marilynn M. Jurovich, Casper, Wyoming
Executive Director-Secretary Fred T. Mahaffey, Chicago, Illinois

Executive Committee
Chairman Carl Shetchek, Eugene, Oregon
Donald Bell, Pittsburgh, Pennsylvania
John P. Plummer, Fairmont, West Virginia
Martin Golden, Dover, Delaware

acceptable scientific testing practices to both report and interpret scores. Perhaps the intention is to emphasize criterion-referenced approaches to score interpretation. If so, care should be taken to specify how criteria are established both from the public policy standpoint and from the standpoint of the current state-of-the-art of the tests and measurements field.

Thank you for the opportunity to comment on H.R. 3564. The language we have criticized would make the bill entirely unworkable and would negate the bill's purpose and overall effect.

Sincerely,

Fred T. Mahaffey, Pharm.D.
Executive Director

FTM/m

cc: NABP Executive Committee

An identical letter was also received from Dr. Mahaffey representing the Federation of Associations of Health Regulatory Boards.

Carnegie Council on Policy Studies
in Higher Education
2150 SHATTUCK AVENUE BERKELEY, CALIFORNIA 94704 (415)849-4474

August 14, 1979

The Honorable Carl D. Perkins
U.S. House of Representatives
2252 Rayburn Building
Washington, D.C.

Dear Carl:

This is to tell you of my concern with bills to regulate standardized testing, H.R. 4949 and H.R. 3564 now before your Subcommittee on Elementary, Secondary and Vocational Education.

Both bills reflect an awareness that standardized tests have come to have an important place in the American educational system. Those who have developed the tests are also acutely aware of this fact and, to my knowledge, have been extraordinarily conscientious in carrying out the responsibilities their role implies. But the very objectivity they have sought to guarantee for the test development and administration process has made standardized testing an almost uniquely impersonal part of the educational system.

This impersonality is now being widely interpreted as a denial of the worth of the many valuable human qualities the tests do not purport to measure. These qualities should be taken into account in an educational admissions program. They are so taken into account wherever the admissions process is conducted in a knowledgeable and far-sighted manner.

The need, however, is for each institution to have such a balanced admissions process, not to discard the objectivity of tests or to make objectivity more difficult to achieve. It is worth emphasizing--and considering in the deliberations of your committee--the great contribution of objective standardized tests in opening up opportunities for many, many students from disadvantaged backgrounds. It is through such tests that such students who have come from unknown schools, have lacked teachers who could give them useful recommendations and who have sought careers contrary to race and sex stereotypes have been enabled to demonstrate their promise and to enter outstanding institutions.

There will be many students in the future, also, who will need such opportunities. They will be ill-served if it becomes harder to take such tests, or more expensive or if so many test questions are placed in the public domain that it will be increasingly difficult to frame questions that genuinely and fairly test knowledge and ability. Standardized tests are valuable servants of the educational system and of the students who have no other way to show that they will make good use of educational opportunities.

Sincerely,

Clark

Clark Kerr

On a personal note I was raised on a farm and went to country grade schools (except the first four grades in a one-room schoolhouse). The tests [illegible] (1) got me into a good college and (2) convinced me that I might be able to [illegible] [illegible].

UNIVERSITY OF VIRGINIA
CHARLOTTESVILLE
22903

OFFICE OF THE
VICE-PRESIDENT AND PROVOST
THE ROTUNDA

TEL. 804-924-[illegible]

August 14, 1979

Mr. Carl D. Perkins
Chairman, Education and Labor Committee
Rayburn House Office Building
Washington, D. C. 20515

Dear Mr. Perkins:

I write to express my concern about the Weiss Bill (HR4949) now under consideration by your subcommittee. As I am sure you are aware, all public colleges and universities are being continually confronted with a demand for accountability, and as a major public institution we at the University of Virginia feel that it is imperative to make the most effective use of the facilities and resources entrusted to us by the Commonwealth. One of our most difficult problems is that of making decisions about admissions of students to the various programs and schools of this institution. We have found that a wide spectrum of information is necessary in making these difficult decisions about what students to admit, and the results of the various standardized tests furnish us some of the most useful information we receive.

Our problem is a huge one. We receive about 25,000 applications for admission to the various parts of the University, and we can admit only about 5,000 new students each year. It is highly important for us to have the best information we can have about the people who apply for admission so that we can best serve Virginia and be accountable to the General Assembly.

Our objections to the Weiss Bill have to do especially with Section 5, which we are certain would have the effect of destroying the quality of the various standardized tests or pricing them so high that the cost would be damaging to the people who take them or both. We think that Section 5 if enacted would in large measure destroy the utility of standardized tests as part of the admissions process.

Mr. Carl D. Perkins
August 14, 1979
Page Two

We have two arguments to make against Section 5. The Weiss Bill is designed, say its proponents, to serve the purposes of students as consumers. We think it is at least equally important that the Congress consider the interest of another group of users, the academic institutions. Further, the history of standardized national tests in this country is enlightening. Before good tests for admissions were developed many colleges and universities, particularly the prestigious private ones, had in large part an admissions process which could legitimately be called "the old boy system." That system was clearly elitist. This historical fact seems to escape proponents of the Bill, who fail to realize that the wide spread use of standardized admissions tests has had a democratizing effect. Indeed, some proponents of the Weiss Bill argue that standardized tests are elitist. They are not elitist; they are meritocratic. There is a vast difference between an elitist society, one based upon inherited privilege, and a meritocratic society, such as Thomas Jefferson envisioned, based upon the demonstrated merit of people.

Sincerely yours,

David A. Shannon
David A. Shannon
Vice President and Provost

DAS:er

EASTERN VIRGINIA MEDICAL AUTHORITY · 1964 ·

EASTERN VIRGINIA MEDICAL SCHOOL
700 OLNEY ROAD
POST OFFICE BOX 1980
NORFOLK, VIRGINIA 23501

OFFICE OF THE DEAN

TELEPHONE (804) 446-5800

August 15, 1979

Honorable Carl D. Perkins
Chairman
House of Representatives
House Office Building
Washington, DC 20515

Dear Mr. Perkins:

We strongly urge your non-support of H.R. 4949.

In our judgment, the impact of this bill, as currently structured, would seriously affect our ability to maintain the integrity, high quality, and usefulness of the New MCAT as one criterion in the highly complex process for evaluation of medical school applicants.

The destructive provision of H.R. 4949 is the requirement that a copy of the New MCAT examination and correct answers must be publicly disclosed after each test administration. The highly specified nature of the New MCAT precludes the development of an unlimited supply of new questions.

This legislation could adversely affect our ability to select candidates capable of attaining high level scientific understanding and clinical performance as future physicians and therefore affecting the quality of health care for the citizens of this country.

Sincerely,

Robert M. McCombs, Ph.D.
Interim Dean

RMM/cdb

NORTHWESTERN UNIVERSITY
CHICAGO, ILLINOIS 60611

Office of the Dean
THE MEDICAL SCHOOL
Ward Memorial Building
303 E. Chicago Avenue
312/649-8186

August 16, 1979

Members of the Education and Labor Committee,
U. S. House of Representatives -
✓Mr. Carl D. Perkins, Chairman
Mr. Paul Simon (Illinois-24)
Mr. John N. Erlenborn (Illinois-14)
Mr. Daniel B. Crane (Illinois-22)
Mr. Robert H. Michel, Minority Whip (Illinois-18)
Washington, D.C. 20025

Gentlemen:

The following letter is written to express my concern about provisions of HR 4949 recently introduced by Mr. Theodore S. Weiss, of New York. My specific concerns are the impact of the bill on qualifying examinations offered through the medium of the Association of American Medical Colleges and the National Board of Medical Examiners.

These examinations are of the highest quality and are among the most carefully prepared and thoroughly researched examinations offered anywhere in the world. The preparation of the examinations is painstaking and accomplished through the services voluntarily offered by hundreds of physicians and scientists throughout the country and in Canada. All examinations used are pre-tested and the results of each examination subjected to computerized scrutiny to eliminate material that is misleading or non-discriminating. The results of the examinations are indispensible to medical schools both in the admissions process to medical schools and in evaluation of the students as they progress through medical education. While the MCAT's are highly predictive in the ability of candidates for admission to satisfactorily pursue and conclude their studies, the NBME examinations provide invaluable assistance in measuring students' ability to become a registered practitioner of medicine. In fact, as you probably know, successfully passing the three parts of National Board examinations is accepted by many states in lieu of another licensing examination. Finally, upon application by the Dean of a medical school, a profile of the performance of his/her students in any given subject can be obtained, thus allowing the Dean and chairman of a department to ascertain the reliability of the curriculum in covering all subject matter as measured against national standards.

Each student learns how he/she performs in each subject. To require the testing agency to provide the performance in each question and the right answer for each question as well as to make a copy of the total examination public, would lead to destruction of both testing agencies. The examinations are extremely complex to prepare, pre-test, research and evaluate. Providing questions and answers would be laborious and costly.

THE McGAW MEDICAL CENTER OF NORTHWESTERN UNIVERSITY

Of greater significance is the fact that the pool of questions available would in a few years become totally exhausted, thus destroying the validity of the use of the examinations. Occasional copies of past examinations have been stolen and are available in Mexico for large sums of money. I mention this only to underline the value of the availability of the written examination.

The proposed regulations can serve only one purpose; to destroy an extremely high grade evaluative process, to further complicate the process of admitting students to medical school and evaluating their progress through the school, and to reduce to mediocrity the methods of determining who is competent to practice medicine.

If I can be of any help in supplying additional information, please let me know.

Sincerely yours,

NORTHWESTERN UNIVERSITY MEDICAL SCHOOL

James E. Eckenhoff

James E. Eckenhoff, M.D.
Dean

JEE:pc

University of Illinois at the Medical Center, Chicago
UNIVERSITY OF ILLINOIS COLLEGE OF MEDICINE
Office of the Executive Dean
1853 West Polk Street, Chicago, Illinois 60612
(312) 996-3500

August 16, 1979

The Honorable Carl D. Perkins
U.S. House of Representatives
Washington, D.C. 20515

Dear Mr. Perkins:

On July 31, 1979, Dr. John A.D. Cooper, President of the Association of American Medical Colleges, testified before the House Subcommittee on Elementary, Secondary, and Vocational Education, Committee on Education and Labor. Dr. Cooper's comments were addressed to two bills, H.R. 3564 and H.R. 4949, currently pending. The former is entitled, "The Truth in Testing Act of 1979," and the latter is entitled, "The Educational Testing Act of 1979."

It is our belief at the University of Illinois College of Medicine that neither of the bills would assist our Admissions Committee in its efforts to use fair and impartial procedures for selecting candidates for admission to the College. We believe that the new Medical College Admission Test (MCAT), when used in the spirit for which it was created, provides us with a valid and reliable measure of our applicants' potential for success in medical school. We further believe that this examination covers only a finite amount of material for which only a finite number of reliable questions may be developed and used. Publication of such examination items would ultimately serve to invalidate the reliability of the new MCAT and cause our Admissions Committee to rely solely on grades (given by faculty and schools of variable quality), letters of recommendation (provided essentially on student request), and a relatively short interview that may assist in identifying non-cognitive variables, but certainly not those of a cognitive nature. Despite the fact that the results of the new MCAT do not constitute the largest portion of our data base, a potential loss of reliability of that examination will make our very difficult job all the more difficult. We would further run the risk of denying admission to otherwise well-qualified students who, by virtue of attending an undergraduate program that does not give grades, can present no cognitive credentials.

We, therefore, strongly support Dr. Cooper's presentation and most respectfully urge that the bills in question be tabled until more information on their own validity and ultimate effects can be obtained and certified.

Thank you in advance for your attention and support.

Sincerely,

Marten M Kernis

Marten M. Kernis, Ph.D
Deputy Executive Dean

THE UNIVERSITY OF ROCHESTER
MEDICAL CENTER
SCHOOL OF MEDICINE AND DENTISTRY · SCHOOL OF NURSING
STRONG MEMORIAL HOSPITAL

601 ELMWOOD AVENUE
ROCHESTER, NEW YORK 14642
AREA CODE 716 275-4022

CENTER FOR BRAIN RESEARCH
BOX 605

August 17, 1979

Congressman Carl Perkins
2328 Rayburn House Office Building
Washington, D.C. 20510

Re: HR 4949

Dear Sir:

I am appalled by the provisions of this bill. The results of the bill will be the elimination of one of the few relatively objective criteria we have for evaluating potential medical students. These tests were developed to avoid the subjective criteria, or at least to serve as a counterbalance to the subjective criteria, obtained in interviews and letters of recommendation. The effect of the bill will be to increase the importance of subjective, potentially prejudiced individual judgements relative to the much more objective and "color blind" data derived from testing.

With rampant grade inflation and tremendous variability in grading criteria from one University to another, the MCAT test has been the only objective criterion for comparison with the subjective judgement. A mixture of objective and subjective criteria are desirable in selecting medical students for a variety of reasons in that some of the personal qualities that are necessary in the physician are not easily detected on testing of this sort. Publishing the questions would greatly increase the expense and difficulty of administering this test and decrease its effectiveness. The expenses involved in developing a new test, with new questions every year, would add greatly to the cost to the student without significantly helping him. In addition, a different test every year would make it impossible to detect trends in performance from year to year.

In summary, I find the bill very mischievous, and for the second time in my life, have been sufficiently bothered by a pending bill to write a Congressman about it. I hope you will oppose this legislation.

Sincerely,

Robert M. Herndon, M.D.
Director, Center for
Brain Research

RMH/nj

UNIVERSITY OF VIRGINIA
FACULTY OF ARTS AND SCIENCES
[illegible] CABELL HALL
CHARLOTTESVILLE [illegible]

OFFICE OF THE DEAN
(Code [illegible]) [illegible]

August 17, 1979

Committee on Education and Labor
Office No. B346C
Rayburn House Office Building
Washington, D.C.

Gentlemen:

I hope very much that you will take a stand against the proposed Weiss Bill, H.R. 4949. Its effect would seem to be that of making standardized testing so difficult to administer that the future of such testing would be in great doubt, were the bill passed. In addition, if it were passed it would be unwarranted interference of the federal government in higher education.

Sincerely yours,

Edwin E. Floyd

Edwin E. Floyd
Dean

EEF/cd

AMERICAN DENTAL HYGIENISTS' ASSOCIATION

WASHINGTON, D.C. OFFICE
1101 17TH STREET N.W.
SUITE 1006
WASHINGTON, D.C. 20036
(202) 833-3024

August 20, 1979

Honorable Carl D. Perkins,
  Chairman
Subcommittee on Elementary,
  Secondary and Vocational Education
2365 Rayburn Building
Washington, D.C. 20515

Dear Mr. Perkins:

The American Dental Hygienists' Association is grateful for the opportunity to submit a statement on H.R. 4949, "Educational Testing Act of 1979", and H.R. 3564 "Truth in Testing Act of 1979" currently being considered by the Subcommittee on Elementary, Secondary, and Vocational Education.

The Association is not in favor of federal legislation at this time to regulate the testing industry. There is no evidence today that such legislation is needed. The Association feels that a study of the results of the National Academy of Sciences and American Council on Education studies on standardized testing which will soon be released should be undertaken prior to the enactment of federal legislation. With New York state passing a similar bill, there also exists an excellent opportunity to learn from their experiences. Much more evidence should be uncovered from currently available sources before action of this scope is undertaken.

If the proposed bill is enacted, the effect on college testing and admissions may be undesirable. Programs that test small numbers of applicants would be severely burdened economically because of the substantial added costs of preparing and shipping new test forms several times a year. In effect, these smaller testing programs would cease, and colleges would have to rely solely on the larger testing agencies which test enough people to offset the added costs. In most cases, then, colleges would have little choice in selecting appropriate tests.

NATIONAL HEADQUARTERS • 211 EAST CHICAGO AVENUE, CHICAGO, ILLINOIS 60611 • (312) 944-7097

Honorable Carl D. Perkins
Page 2
August 20, 1979

The Dental Hygiene Aptitude Test is the only currently available instrument that is specifically designed for dental hygiene applicants. This test provides information to the applicant on his or her standing in comparison to all applicants, and on his or her standing compared to others with comparable education. Evidence shows that this type of information is useful to admissions officers.

If this proposed legislation is enacted, the Dental Hygiene Aptitude Test would have to be abandoned since the cost would become so prohibitive as to diminish its desirability to institutions and to prospective dental hygiene students. To preserve the integrity of the test, at least three test forms would be required annually. Costs would triple in order to: maintain a suitable size item pool; cover costs for extra printing; ship additional tests; and construct new versions. These costs, of course, would have to be passed along to the consumer.

The DHAT scores are not the only criteria used by institutions in admissions screening. Personal interviews, grade point average, other test scores, and life experiences are also part of the selection process. The DHAT can offer an applicant who has long been out of school information on the likelihood of success; it can also identify individuals who may require remediation. It is obvious that the DHAT can be a useful admissions tool.

The Association firmly believes that any legislation at this point would be detrimental to the DHAT and to the colleges that use this instrument as an initial screening device. The gain in predicting academic success, as compared to selection of students at random, is large enough to warrant its use. At this point the Association urges the adoption of a "wait and learn" attitude to avoid the consequences of a hasty decision that may not be in the consumer's best interest.

Sincerely,

Kathleen A. Mast, RDH

Kathleen A. Mast, RDH
President, ADHA

cc: Subcommittee on Elementary,
Secondary and Vocational Education
ADHA Board of Trustees

COUNCIL FOR BASIC EDUCATION
725 FIFTEENTH STREET, N.W., WASHINGTON, D.C. 20005 (202) 347-4171

COUNCIL FOR BASIC EDUCATION
CBE
FOUNDED IN 1956

OFFICERS AND DIRECTORS

President
CARL J. DOLCE
Dean, School of Education,
North Carolina State University

Vice President
THOMAS C. MENDENHALL
President (1959-1975),
Smith College

Treasurer
EVERETT N. BELLOWS
Vice President,
Olin Corporation

JACQUES BARZUN
The University Professor Emeritus,
Columbia University

MRS. TALCOTT BATES
Former member, California
State Board of Education

MRS. BARRY BINGHAM
Vice President of the Louisville
Courier-Journal and Times

HENRY N. DREWRY
Director, Teacher Preparation and
Placement, Princeton University

CLIFTON FADIMAN
Author and Critic

FRANCIS KEPPEL
Director, Education Program, Aspen
Institute for Humanistic Studies

WILLIAM H. MARMION
Employee Relations Officer,
Long Beach, California, Schools

MRS. EDWARD B. McMENAMIN
Headmistress,
The Nightingale-Bamford School

[illegible]
Director, Planning & Development,
University of Dallas

THOMAS W. PAYZANT
Superintendent of Schools,
Oklahoma City

MORTIMER SMITH
Former Executive Director
Council for Basic Education

MICHAEL J. SPECTOR
Attorney
Quarles & Brady, Milwaukee

STAFF

A. GRAHAM DOWN
Executive Director

H. DENNIS GRAY
Assistant Director

JAMES HOWARD
Assistant Director

[illegible]
Program Assistant

[illegible]
Program Assistant

PAUL KENNEDY
[illegible]

August 29, 1979

Honorable Carl D. Perkins
Committee on Education and Labor
Rayburn House Office Building
Washington, DC 20512

Dear Congressman Perkins:

The Council for Basic Education has a long-standing interest in standardized testing in the nation's public schools. Your Subcommittee on Elementary, Secondary and Vocational Education is currently considering two proposed bills, H. R. 3564 and H. R. 4949, on which the Council wishes to comment briefly. For a more detailed analysis of college admission tests, I refer you to the enclosed 1974 publication of the Council, Uses and Abuses of Standardized Testing in the Schools by George Weber (pages 22-33). I have also enclosed some information about CBE.

As far as college admission tests are concerned, there is no need for either of the proposed bills. Both imply causes for legislation that simply do not exist. Contrary to these implications, college admission testing provides a model for all test makers to use in explaining their tests to test subjects and test users, the people most concerned. (See Weber, pp. 25-28.) Unfortunately, journalists, legislators, and the public sometimes blur the significant distinctions that properly should be drawn between college admission tests and other kinds of standardized tests that are indeed flawed. The Congress probably cannot legislate successfully against such ignorance.

It may be that college admission testing has been damned by its mistaken association with bad or abused tests. Group "intelligence" tests, for example, probably should be abolished, because they stigmatize children unfairly and because they do not actually measure what they purport to measure--native mental capacity. (See Weber, pp. 3-6.) Misuses of good tests can be cited as well. The results of standardized achievement tests are widely misinterpreted by teachers, school boards, legislators, journalists, and the public. (See Weber, pp. 13-16). In all forms of testing, abuses are possible; in most they occur frequently.

August 29, 1979
Page Two
Honorable Carl D. Perkins

But college admission tests rate high marks by any standard of judgment. Perhaps this is because the existence and use of college admission tests are determined outside the schools whose students are being tested, a fact that means reduced likelihood of abuse.

With no help from government, college admission tests already effectively accomplish their intended purposes for colleges and students without harm to either. These tests are not afflicted with ills that require new legislative cures, nor do the two proposed bills offer any improvement in the admission tests or the testing process we now have. In short, the bills appear to be unneeded efforts to deal with non-existent problems--hardly solid ground for federal action.

Should you wish me to develop these ideas in greater detail as testimony for the Subcommittee, I would be pleased to do so.

Sincerely,

A. Graham Down
Executive Director

AGD/srb
encls.

cc: Members of Committee on
Education and Labor

SEP 18 1979

MILLSAPS COLLEGE Jackson, Mississippi 39210
OFFICE OF THE PRESIDENT

September 7, 1979

George M. Harmon

The Honorable John Hinson
1512 Longworth
House Office Building
Washington, D. C. 20515

Dear Mr. Hinson:

It has come to my attention that you will be sitting on a committee considering two Bills: The Educational Testing Act of 1979 (HR 4947) sponsored by Congressman Wise; and The Truth and Testing Act of 1979 (HR 3564) sponsored by Congressman Gibbons. In my judgment these Bills would have a severe debilitating effect on standardized testing as it is used constructively by educational institutions to assist and provide guidance to students.

Standardized tests, while not the only criteria, are invaluable as tools for educators in assisting students to understand themselves and to make wise educational and career choices. Laws violating the security of these instruments will significantly increase the cost of their development, thus limiting their availability to the disadvantaged. Because of limited items available to some testing areas the possibility of impartial testing may be removed entirely. Further, the door will be opened for "test taking" courses based on the items made public, again creating an unfair advantage for the affluent and possibly leading to incorrect decisions relative to the uses of the test.

It is my hope that you will oppose and work for the defeat of both Bills. If you have questions I will be more than glad to provide further information at least as regards this institution and its concern.

Sincerely,

George M. Harmon
President

GMH/mn

STATE OF IOWA

Commissioners
Gracie M. Ruden, LeMars, Chairperson
John J. Poggs, Council Bluffs, Vice-Chairperson
Julian C. Campbell, Bloomfield
Mildred E. Elliott, Mount Ayr
Monroe J. Colston, Des Moines

C. R. GALVIN
DIRECTOR
TELEPHONE (515) 281-3183

MEMBER
NATIONAL ASSOCIATION OF
REAL ESTATE LICENSE LAW OFFICIALS

REAL ESTATE COMMISSION
1223 E. Court Ave., Des Moines, IA 50319

September 10, 1979

Chairman Carl Perkins
Subcommittee on Elementary, Secondary,
and Vocational Education
Committee on Education and Labor
B-346-C Rayburn House Office Building
U.S. House of Representatives
Washington, D.C. 20515

RE: H.R. 3564
Weiss and Gibbons Bills

The Iowa Real Estate Commission can well appreciate the Legislators thinking that there is a possibility of a need for some federal regulation governing licensing and certification, but the Iowa Real Estate Commission is of the strong belief that this type of regulation can be regulated more efficiently on a state level without federal legislation.

In Iowa each applicant for a real estate license must pass a written examination authorized by the Commission and administered by the Commission or persons designated by the Commission. The examination shall be of scope and wording sufficient in the judgment of the Commission to establish the competency of the applicant to act as a real estate broker, salesperson, or apprentice salesperson in such manner as to protect the interest of the public.

The Iowa Real Estate Commission urges the Committee on Education and Labor in its investigation of the regulations governing occupational admission tests to recognize that the regulations of such business has been actively delegated to the Iowa Real Estate Commission by the state legislature since 1930. License laws have been and are regularly being revised to meet the needs brought about by changes.

Chairman Carl Perkins
September 10, 1979
Page Two

There has been close cooperation with the Civil Rights Commission or committee to see that the contents and the structure of the examination does not discriminate.

The Iowa Real Estate Commission does not believe that there is a convincing case for federal legislation at this juncture, and that the occupational admission test should remain the obligation of each state and not a federal regulated program.

C. R. Galvin

C. R. Galvin
Director

CRG/cw

SEP 18 1979

THE UNIVERSITY OF MISSISSIPPI MEDICAL CENTER
25[illegible] North State Street
JACKSON, MISSISSIPPI 39216

School of Dentistry
Office of the Dean

Area Code 601
968-48[illegible]

September 12, 1979

The Honorable Jon C. Hinson
1512 Longworth
House Office Building
Washington, D.C. 20215

Dear Representative Hinson:

I am very concerned about HR 4949, the Educational Testing Act of 1979 introduced by the Honorable T. Weiss. Having reviewed the bill and the remarks about it in the Congressional Record dated July 25, 1979 I am disturbed over the possible implications for dental education.

I have discussed the proposed bill with my chief administrators and certain members of the faculty of the School of Dentistry. It is my understanding that Bill 4949 may be reported out by the subcommittee in late September and may be discussed by the House Committee on Education and Labor soon after. Since you are a member of this important committee, I want to express my concern over the possible effects such legislation would have on our new and rapidly developing School of Dentistry in Jackson. Listed below are the salient points from our joint discussions which I would like to bring to your attention.

-Generally the bill intends to alleviate bias in testing, improve public accountability and mandate financial disclosure. These are notable but we fail to see how this bill will alleviate bias in testing and improve public accountability. It can, we believe, mandate financial disclosure.

-In relationship to test bias, the bill cites studies relating income to test scores. One could just as well cite studies showing that there is a moderate correlation between intelligence level and income. Perhaps such a correlation between income level and test scores reflects intellectual factors? In other words, the problem of test bias is much more complex than that which is addressed in the bill.

-Test makers for years have attempted to develop tests that would have predictive validity (i.e. predict criteria with a high correlation between test scores and the criteria). Research would support the fact that many of the aptitude testing programs in fact do a fairly good job of this thus saving the student and taxpayer much frustration and concern. The background data presented in this bill is biased from this standpoint when they say correlations are very low.

The Honorable Jon C. Hinson
September 12, 1979
Page 2

-We believe the bill is too restrictive and places unfair guidelines on the test companies, especially the requirement for companies to send a copy of the question and answer sheets to the subject. This is unrealistic and in some testing areas item pools would be depleted very quickly. Since all possible items would be available to anyone, it would defeat the purpose of testing.

-Some points of the bill such as providing more information to the subject concerning the purpose of the test, skills demanded, reporting of scores and privacy of students are necessary. If testing companies are in fact taking advantage of the public related to costs, then the accountability section of the bill would seem to be appropriate. However, very little background data is provided on financial accountability.

In summary, this bill does not seem to be realistic in what it hopes to accomplish and the methods of doing so are impractical. We believe a more adequate bill with much less restriction should be written. I would very much appreciate your reaction to the concerns of the administration and faculty of our School of Dentistry and please accept my appreciation for your consideration.

Sincerely,

Wallace V. Mann Jr.

Wallace V. Mann, Jr., D.M.D.
Dean, School of Dentistry

WVM/ajt

cc: Dr. Harry Bruce, Executive Director - American Association of Dental Schools
Dr. Thomas Ginley, Secretary - Council on Dental Education
Dr. James Graham - Assistant Secretary - Council on Dental Education

Mississippi College
OFFICE OF THE PRESIDENT
CLINTON, MISSISSIPPI 39058
TELEPHONE 601-924-5131

September 13, 1979

Honorable Jon Hinson
US House of Representatives
Washington, DC 20215

Dear Jon:

I wanted to write you with regard to two related bills (H.R.3564 and 949) which are currently under study by the House Committee on Education and Labor, of which you are a member, I believe.

I simply wanted to express my very strong opposition to this type of legislation; once again, it's like taking a cannon to kill a fly! This is legislation somewhat similar to that which has passed in New York state and is being considered in a number of other states. Such legislation strikes perilously close to the heart of the academic enterprise. Such legislation would have the general effect of requiring that test questions and answers be disclosed to students after their test results are returned as well as requiring that certain information be distributed to students at the time they register to take the tests (ACT, SAT, MCAT, LSAT, etc.). Also, they require that all studies involving test results be filed as a matter of public record. Further, one of the proposed pieces of legislation would limit the use of test results in selection decisions.

Basically, I am opposed to such legislation because it is quite apparent that Representative Weiss from New York is trying to place this on the federal level since MCAT opted out of New York state after the New York legislature approved a similar bill. If that's what New York wants, that is their privilege but I do not believe that is necessarily good for the nation just because New York may desire it.

Althought the proponents cite admirable goals and disclain any intention to interfer with the admissions processor with the nature of the tests, the legislation would accomplish just these goals in a most ineffective way. Further, the potential impact on institutions as well as students and their families has not been clearly established nor have the relative costs and benefits to students and society been documented. Finally, the limited evidence thus far presented does not seem to document sufficient pervasive harmful results of test misuse to justify legislation intervention nor to assure that the cure would not be worse than the disease.

I shall appreciate your thoughtful consideration of these matters and look forward to being in Washington in the not too distant future to discuss these as well as other issue with you.

With very best wishes,

Lewis

Lewis Nobles, President

LN:ps

STUDENT
FUTURES&
PLACEMENT

Carleton College
Northfield, Minnesota 55057
[illegible] 507/663-4294

September 14, 1979

Congressman Arlen Erdahl
House of Representatives
Washington, DC 20515

Dear Congressman Erdahl:

I am aware of the "Truth in Testing" legislation which is currently being debated in Congress, referred to the Education and Labor Committee.

As a mid-level administrator at Carleton College for 17 years, I have counseled students and alumni regarding tests required for admission to graduate and professional schools, and I have also administered many tests which would be affected by the pending legislation.

In my opinion, "Truth in Testing" provisions would be harmful for these reasons:

(1) The tests would become more expensive, since new questions would have to be written for each new administration. A valid test is a very valuable document, and you cannot imagine the great care that is given to writing, validating, "de-culturing" a test. After that, extreme security is practiced in the distribution and collection of tests. When one considers the labor which is given to that effort, the costs of testing borne by the test takers is very low. The testing business is an extremely efficient one, I feel, and I've worked with them a long time. By the way, I do not consider myself an advocate of testing companies--my allegience is to my job as a placement professional--my first loyalties are to students who take the tests. I feel they are served very well.

(2) Pressure would build for test preparation. Expensive workshops to "cram" for law, medical and other boards would proliferate. More expense for the serious applicant! But my concern is also that tests are just one of many considerations that a college/ university should consider in admitting a student. I think this is practiced now in most cases. Federal legislation which would enable the wealthy or the conniving type of individual to enhance his chances of college admission is contrary to enlightened practice. Present policy serves very well--practice questions and general information about test questions are available to all.

(3) The quality of testing would have to sink if security was broken each time tests are administered. Tests in all reputable companies are prepared with the help of consultants and committees from their users. Members of Carleton's staff help in the work occasionally. Coming from a wide variety of backgrounds, they insure that tests are written which insure, as much as possible, the broad representation of questions desired for the test's purpose. These people, the test companies could not keep up the needed quality of work in the fact of constant need for new tests, unless there were huge amounts of money expended. More expense! Greater margin for error! It is possible that testing quality would drop so that testing would lose credibility to the extent that it would be unusable. That would put us back in the Dark Ages

(continued)

again as far as the selection of candidates for educational opportunity is concerned.

May I say, in closing, a few things about the performance of test companies which I'm most familiar with. They are the Educational Testing Service of Princeton, NJ, the American College Testing Company, Iowa City, and the Psychological Corporation, New York. Their delivery of services is exemplary. They have earned and enjoy a high level of trust in the academic community, including the overwhelming number of students, for validity, efficiency, and, beyond that, an enviable performance in promoting equal opportunity to the general public.

I would hope that you will consider carefully the "Truth in Testing" legislation. I feel it is unneeded and counterproductive to good educational practice.

Sincerely,

Marilyn J. Hill

Marilyn J. Hill
Associate Director
Placement Coordinator

EDWARD R. MADIGAN
ILLINOIS

Congress of the United States
House of Representatives
Washington, D.C. 20515

September 19, 1979

The Honorable Carl Perkins
Chairman
Subcommittee on Elementary, Secondary
and Vocational Education
B 346C

Dear Carl:

The enclosed correspondence from the president of Illinois State University, Dr. Lloyd Watkins, addresses the issues involved in two truth in testing bills currently under consideration by your Subcommittee on Elementary, Secondary and Vocational Education. I hope President Watkins' comments will be useful in your deliberations on this legislation.

Sincerely,

Ed

Edward Madigan
Representative in Congress

EM/jad

THIS STATIONERY PRINTED ON PAPER MADE WITH RECYCLED FIBERS

Illinois State University
President

September 6, 1979

The Honorable Edward R. Madigan
U. S. House of Representatives
1514 Longworth House Office Building
Washington, D. C. 20515

Dear Representative Madigan:

My purpose in writing is to oppose the proposed legislation of The Weiss bill (H.R. 4949) and The Gibbons bill (H.R. 3564), the so-called "Truth in Testing" legislation and similar to The New York State bill (S. 5200, LaValle et al.) which was recently enacted.

In opposing this proposed legislation I am not unaware that some misuse of standardized testing does occur. The solution to the problem of improper use of standardized tests, however, is not to enact legislation as proposed in the Weiss and Gibbons bills. The disadvantages of these two bills far outweigh the advantages. The disadvantages are several, and each is substantive as a potential disservice to graduate education in the United States. First, and primary, is the threat to test quality. It may not be possible to maintain the present quality and value of the test as an aid for decision-making for student admissions. This is particularly true for the immediate future when the requirement for disclosure of test questions and answers will result in compromised tests.

Second, the proposed legislation, if enacted, would increase student costs. The constant development of new tests, which is extremely difficult, is expensive. Third, the proposed legislation would probably result in fewer test administrations, thereby decreasing access to taking the test for the student. This could have an adverse effect on admissions processes.

Your consideration of these disadvantages in the hearings on October 10 and 24 will be appreciated. In my opinion, the legislation should not be enacted.

Sincerely,

Lloyd I. Watkins
President

LIW/gc

Normal-Bloomington, Illinois
Phone: 309/438-5677

Hovey 308
Normal, Illinois 61761

Equal Opportunity/Affirmative Action University

AMERICAN BOARD OF MEDICAL SPECIALTIES

GLEN R. LEYMASTER, M.D.
Executive Director

JAMES L. HANSEN, M.D.
Associate Director

1603 Orrington Avenue, Suite 1160
Evanston, Illinois 60201
(312) 491-9091

October 9, 1979

The Honorable Carl D. Perkins
Chairman, Subcommittee on Elementary,
Secondary and Vocational Education
Room B346C
Rayburn House Office Building
Washington, DC 20515

ATTENTION: Nancy Cober

Dear Sir:

This letter transmits to you testimony which the American Board of Medical Specialties had hoped to be able to present in person to the Subcommittee on Elementary, Secondary and Vocational Education since it is directed to two bills, HR-3564 and HR-4949, which are currently being examined by your Committee.

Since we shall not have the opportunity to testify in person before your Committee, I am submitting written material with the request that it be considered by your Committee as part of the deliberations on the "truth in testing" legislation, in accord with the suggestion of your staff.

It seems appropriate to me to submit the statement as we have prepared it for oral presentation; I trust that your Committee will also consider this form satisfactory. If there are any questions regarding the testimony, the position of the American Board of Medical Specialties and any of its member boards in regard to this legislation, I shall be pleased to respond on behalf of ABMS or to secure authoritative responses from our member boards, whichever procedure might best secure the information which you desire.

If additional hearings are scheduled at a future date, and if you deem it appropriate for a representative of the American Board of Medical Specialties to present testimony, please let me know and we shall be pleased to send an informed representative.

Yours sincerely,

Glen R. Leymaster, M.D.
Executive Director

GRL/jem

Attachment: American Board of Medical Specialties
Testimony Regarding Proposed
"Truth in Testing" Legislation

Testimony Regarding Proposed "Truth in Testing" Legislation before the House Subcommittee on Elementary, Secondary, and Vocational Education of the Committee on Education and Labor; Hon. Carl D. Perkins, Chairman.

Good day, Mr. Chairman, and members of the Committee. The American Board of Medical Specialties appreciates this opportunity to present its views on HR 3564 and HR 4949.

[Description of ABMS]

I am Dr. Gordon W. Douglas, President of the American Board of Medical Specialties. Accompanying me today is Dr. Glen R. Leymaster, Executive Director and Secretary of the American Board of Medical Specialties. The American Board of Medical Specialties, which we refer to as "ABMS," is an organization whose purpose is to assist the 23 medical specialty boards which are members in their individual efforts to evaluate and certify physicians specializing in the various fields of medicine. The 23 member specialty boards each conduct voluntary certification programs in the field of medical practice they represent. These specialties range from anesthesiology to urology. I would emphasize that these certification programs are voluntary. They do not affect a physician's license to practice; rather certification by a medical specialty board constitutes formal recognition of a physician's competence in a chosen specialty by his or her professional peers.

As an umbrella organization, ABMS also serves as a spokesman for and representative of its member specialty boards within the medical profession. For example, ABMS was one of the organizers of such bodies as the Coordinating Council on Medical Education, the Liaison Committee on Graduate Medical Education, and the Liaison Committee on Continuing Medical Education. Thus, ABMS is closely concerned with medical education and is involved in a variety of efforts to assure that a high level of physician competency is maintained in the medical profession. As a result, I expect that my testimony will reflect many of the same concerns expressed by representatives of the National Board of Medical Examiners, the American Association of Medical Colleges, the Federation of State Licensing Boards, and others. I believe it is fair to say that virtually all of these professional groups have grave concerns about the effect the proposed legislation might ultimately have on the standard of medical practice in the United States.

[Problems in the Proposed Legislation]

While it does not appear that the bills presently before this committee would directly affect the voluntary testing and certification programs of the various medical specialty boards, ABMS is concerned that these bills may mark the beginning of a development that would ultimately be detrimental to the goal of assuring the highest possible level of professional competency among physicians.

Of course, ABMS has no quarrel with the stated goals of the proposed legislation. Everyone can agree that tests used in evaluating educational admissions or entrance to various occupations should be fairly and equally administered. That is not the only worthwhile goal to be pursued, however. Particularly in the case of highly sophisticated professions like medicine, the public has an overriding concern that a high level of individual competence and excellence be maintained. Indeed, the medical profession has always approached the evaluation of physician competency as a public trust. Because of the obvious adverse effects of allowing any but the most skilled individuals to practice medicine, the medical profession has always sought the best possible testing programs to insure the excellence of all who are admitted to practice.

In a very real sense, medical testing programs -- and testing programs developed by the other learned professions like dentistry and law -- have a fundamentally different purpose than tests used in other vocational settings. In many occupational areas, tests merely serve to verify that an applicant has some minimal level of skills. In these settings, public accountability for the fair and even-handed administration of the test may be of tantamount concern. By contrast, professional testing programs strive to assure that professional practitioners demonstrate the high levels of competence, skill, and knowledge commensurate with the profession's public trust. In this context, the foremost test criteria must be the assurance that the test used will successfully screen out all but the genuinely competent applicants. Thus, appropriate testing standards must be geared to the testing context. We submit that inflexible standards of test administration, such as those found in the present bills, should not be allowed to undercut the assurance of professional medical competency which the American public expects and deserves.

[Disclosure of Questions and Answers]

I am sure that the committee has already received testimony from a variety of individuals and groups, including other spokesmen from the medical profession, regarding some of the specific problems and hazards posed by the proposed bills. I will briefly touch upon one item which appears to be crucial to a wide variety of testing programs now used by the medical profession.

The most troublesome provision in the proposed legislation is the requirement that test questions and correct answers be disclosed. This poses two problems. First, once published, we believe it would be the height of irresponsibility to reuse a test question. We do not want physicians to be licensed or certified because they can learn correct answers from past examinations; we only want physicians to be licensed if they have a professionally sound knowledge of medicine. In some highly specialized areas of medicine, the need to constantly devise new test questions may pose practical difficulties.

Second, and more fundamentally, the need to constantly use new test questions may reduce the validity and reliability of the tests themselves. I think everyone would agree that questions which are vague or which contain some hidden flaw that misleads the test-taker are bad questions. Similarly, and particularly in professional testing, we do not want questions that lend themselves to large numbers of correct guesses; rather, we want questions that accurately and reliably measure the test-taker's knowledge. Modern test-construction procedures can eliminate many of these types of bad questions. For example, by statistical analysis, it is possible to weed out questions for which test-takers' answers do not correlate with their overall score. A lack of correlation between how well test-takers do on a particular question and how they do on the overall examination suggests that question really isn't measuring the level of the test-takers' knowledge. Similarly, questions that are generally too hard or too easy can be identified statistically and eliminated.

But these modern test-construction techniques require good information on the past performance of the question in previous examination sessions. By preventing reuse of questions, disclosure of test questions will severly limit the use of sound test-construction methods. In short, the disclosure proposal appears to be counterproductive to the goal of the proposed legislation. It will not make tests fairer -- it will only make them less reliable.

The damage that would be done to medical testing by disclosure of questions and answers would not be offset by any positive benefits. I understand that some previous witnesses before this committee have suggested that disclosure would equalize the advantages now supposedly enjoyed by test-takers who take test preparation courses or that revealing correct answers would be a "learning experience" for the test-taker.

Neither of those supposed benefits has any relevancy to professional testing. The medical profession wants physicians to learn medicine in medical school; not by repeated efforts to simply pass examinations.

[Interference with Professional Standards]

What disclosure may do, however, is to expose educational and professional standards to interference by a wide variety of interest groups who seek not to assure the highest levels of professional qualifications, but rather to assure their own success in gaining entry to a chosen school or field. One can reasonably predict that professions or institutions which attempt to maintain standards of excellence will find themselves embroiled in controversy and litigation, as a result.

More broadly, ABMS is concerned about where the whole concept of a federal standard of "truth in testing" is leading. Regulation of the professions is a role traditionally occupied by the states. The states, in turn, have relied heavily on the various professions to define, improve, and maintain professional standards and qualifications for practice. We are confident that Congress does not intend that this traditional structure of professional regulation should be altered simply to facilitate the prescription of uniform testing standards. Yet, the proposed bills pose that threat. The proposed bills define their scope of coverage in terms of "interstate commerce." I am not a lawyer and I cannot speak to the question of whether medicine is an aspect of interstate commerce as a legal matter.

I am told that some courts have given that concept rather broad meaning in the past. But I can say that the medical profession does not view itself that way. And traditionally the states have not viewed medicine that way, either.

[Exemption for Professional Testing Programs]

In sum, ABMS is opposed to the present bills. If the committee believes that some legislation is required in this area, ABMS urges that the legislation should be carefully tailored to exempt the sophisticated testing programs employed in professions such as medicine. To the best of our knowledge, no one is claiming that there are any significant abuses in medical professional testing programs. None have been brought to our attention. I believe that any review of the present structure of medical testing would show that medicine and a variety of other learned professions are light years ahead of the run-of-the-mill testing methods used in many occupational settings. In view of the serious problems posed by the proposed legislation, its scope should be no broader than absolutely necessary to meet the specific abuses the committee identifies. Medical testing programs should be explicitly exempted.

Thank you.

AMERICAN DENTAL ASSOCIATION

WASHINGTON OFFICE • SUITE 1004 / 1101—17TH STREET, N.W. • WASHINGTON, D.C. 20036 • PHONE 202 833 3036

October 11, 1979

The Honorable Carl D. Perkins
Chairman
Subcommittee on Elementary, Secondary
and Vocational Education
2328 Rayburn House Office Building
Washington, D.C. 20515

Dear Chairman Perkins:

I am writing on behalf of the American Dental Association to express our concerns regarding the disclosure provisions of the Educational Testing Act (H.R. 4949), and related measures, now pending before the Subcommittee.

As you undoubtedly appreciate, the ability to perform highly precise manual procedures is a fundamental skill necessary for the practice of dentistry. Unfortunately, not everyone who seeks admission to dental school possesses this manual dexterity. Until 1950, the national attrition rate in dental schools averaged between 20 and 30 percent. This resulted, in large measure, from the belated discovery that many students entering their clinical years of training were simply unable to master the fine manual techniques required for graduation. In response to this problem the American Dental Association developed what is known as the Dental Admission Testing Program (DAT). Following the introduction of the DAT and its acceptance as a condition for entrance to all accredited dental schools the attrition rate was reduced to approximately 7 percent. A critical element in this demonstrated success of the DAT is a series of perceptual and space relation tests which are designed to measure an applicant's manual dexterity. The current format of this aspect of the examination is nonverbal, consisting of two and three dimensional objects which the test subject must relate to.

The dilemma which is posed by the disclosure provisions of the legislation centers on the fact that there is a finite number of geometric forms which can be represented in a test of this nature. If the Dental Admission Testing Program were required to reveal the actual tests after each examination, as is proposed in H.R. 4949, the dimensional possibilities for measuring perceptual motor skills will have been exhausted within a very short period. When that occurs, these important manual dexterity tests will have to be discontinued because their value will have been compromised.

In summary, the DAT is unique because it is the only entrance examination program which contains tests that measure abilities as well as academic knowledge. A disclosure of the verbal aspects of the DAT, such as its science or reading comprehension tests, could be compensated for in part by other criteria. However, there is no comparable instrument to replace the manual dexterity component of the DAT. Its loss, through disclosure, would severely restrict the ability of the dental schools to discriminate manual technique skills among their applicants.

For that reason we strongly urge the Subcommittee to ensure that any legislation, which may contain test disclosure requirements, provide an exemption for the Dental Admission Testing Program. Such an exemption could, for purposes of the DAT, be accomplished by the inclusion of language stipulating that the disclosure provisions shall not apply to "any perceptual ability test or perceptual ability portion of a larger test" or, "to any standardized test which will be administered to fewer than 10,000 test subjects nationally per administration." I am enclosing a copy of the Association's testimony on this issue which will provide additional background on the DAT. Your consideration of this request is greatly appreciated.

Sincerely,

William E. Allen, D.D.S.
Chairman
Council on Legislation

WEA:vg

Enclosure

bcc: Mr. John F. Jennings

THE JOHNS HOPKINS UNIVERSITY • BALTIMORE, MARYLAND 21218

DEPARTMENT OF PSYCHOLOGY

October 17, 1979

Congress of the United States
House of Representatives
Committee on Education and Labor
Subcommittee on Elementary, Secondary
and Vocational Education
B-346C Rayburn House Office Building
Washington, D.C. 20515

Statement on HR 3564 and HR 4949 Concerning Standardized Testing

Thank you for the opportunity to comment on two bills on standardized testing, HR 3564 and HR 4949. The two bills have the laudible intent of protecting consumers, mainly students and job seekers, from testing abuses. However, some of the mechanisms proposed to provide this protection will do more harm than good.

The major difficulty with HR 4949 is Section 5, which requires that the actual content of each test be disclosed. The apparent intent of this provision is to help the student determine whether his answer sheet was correctly scored. Merely providing an answer key and a copy of the student's answer sheet will permit checking for clerical errors. Publishing the questions will verify the answer key, but has many adverse effects.

Students seeking to improve their scores will argue that no single correct answer exists for most test items. Scoring the items will be challenged in advocacy proceedings. The bill thus invites the substitution of judicial judgment for educational judgment.

Publication of test items will put a tremendous additional burden on those who prepare test items, since many more will be needed. Since every item must be carefully reviewed, pretested, and analyzed statistically before being used operationally, the procedure will be hard-pressed and reducation in quality of items seems inevitable. There is also more chance for biassed items to slip through the screening process.

There is a limit on the number of new questions that can be produced. It will not be possible to provide a new form for every occasion on which tests are given and so services will suffer. Special administrations will no longer be possible for the handicapped, for those who cannot, for religious reasons, take tests on Saturday, and for those needing to be tested before the next available regular date.

"Open" testing works against new kinds of ability measures. Proponents of the bill argue that the legislation will encourage alternative measures, but they surely do not mean alternative tests. For example, the new "analytic aptitude" test in the Graduate Record Examination is still under review and modification two years after introducing it. New types of items are being tried. If all items must be published, a huge reservoir of available items is required before a new test can be introduced, and the tendency will be to avoid anything new. The GRE "analytic" measure, for example, may well have to be scrapped.

Two aspects of HR 3564 need repair, because they advocate questionable test practices. Section 6b3 requires publication of the score required for admission to an occupation or a university. Usually there is no such score, and proper test use avoids specifying a particular score cut-off. Test scores should be used in conjunction with records of past accomplishments, letters of references, interviews and other pertinent information in making a final judgment.

Section 6c of HR 3564 requires that any test of knowledge or achievement as distinct from aptitude, be graded absolutely, rather than relative to other students. Experts differ on how to do this, on whether it can be done, and on how to distinguish aptitude from achievement. The requirement involves technical matters that are still unresolved. It would be premature to "solve" the problem legislatively.

The bills, stripped of the sections mentioned, endorse the test standards already put in place by the American Psychological Association and the American Educational Research Association. They require publication of information already in the public domain. They also anticipate as yet undisclosed recommendations of the distinguished National Research Council committee on testing. At the least, the bills are premature; I believe that they are unnecessary.

Some of the language in the bills suggests that the tests may be biassed in favor of students from families with high income, and, by implication, in favor of whites. Because aptitude tests measure general abilities developed over many years in educational environments, they reflect the quality of education experienced by the student. Average differences in test scores, between rich and poor, white and black, reflect average differences in educational opportunity. The only way to avoid these differences is to improve equality of educational opportunity. Destroying the tests won't help. Also it is shortsighted to focus on the average values. The tests provide an opportunity for each individual to demonstrate their talent. Many fine students come from unknown schools; without tests to demonstrate their skills, they would be considered bad risks.

I see no important benefit from the legislation, and great potential harm. I recommend that the committee decide against the bills.

Respectfully submitted,

Bert F. Green
Professor of Psychology
The Johns Hopkins University

BFG/bl

The Council of Graduate Schools in the United States

One Dupont Circle, N.W.
Washington, D.C. 20036
Phone: 202/223-3791

October 19, 1979

Honorable Carl D. Perkins, Chairman
Subcommittee on Elementary, Secondary, and
  Vocational Education
Rayburn House Office Building, B-346C
Washington, D. C. 20515

Dear Mr. Perkins:

This is in response to your October 9, 1979, letter stating that you will accept a written statement from The Council of Graduate Schools for inclusion in the hearing record concerning the two bills dealing with standardized testing, H.R.3564 and H.R.4949.

Let the record note that The Council of Graduate Schools in the United States, representing the wishes of its graduate deans, is against the passage of H.R.3564 and H.R.4949. Attached in support of this statement is a sampling of letters from our constituents. Some have written directly to you. There are approximately 365 member institutions in The Council of Graduate Schools. We are against H.R.3564 and H.R.4949 for the following reasons:

(1) We believe that the bills are overly precipitous. Before radical change is mandated, careful study should be made of the arguments both for and against the proposed legislation determining the justification and the probable ramifications of such legislation;

(2) We fear that test quality will deteriorate since it may not be possible to maintain the present quality and value of the test as an aid for decision making for student admissions if disclosure of test questions and answers is mandated;

(3) We fear that the student will suffer from increased costs if the testing community is forced to continually replace compromised tests;

(4) We fear that the proposed legislation will result in fewer test administrations, thereby decreasing access to taking the test for the student. In short, a decrease of service to the student.

Thank you for giving The Council of Graduate Schools the opportunity to submit to the record the views of its members.

Cordially,

Michael J. Pelczar Jr.

Michael J. Pelczar, Jr.
President

lrf

Enclosures

cc Webb [illegible]

UNIVERSITY OF MIAMI
CORAL GABLES, FLORIDA 33124

September 19, 1979

Tel. (305) 284-4154

GRADUATE SCHOOL
P. O. BOX 248125

The Honorable Edward J. Stack
U. S. House of Representatives
Washington, D.C. 20515

Dear Mr. Stack:

I am writing to express my concerns about the so-called "Truth in Testing" bills; namely, the Weiss Bill (H.R. 4949) and the Gibbons Bill (H.R. 3564). My major concern is with the Weiss Bill, specifically the requirements for disclosure of all questions on the test together with the correct answers.

Much time and money is dedicated to the construction of these tests to insure their comparability at different times and in different places and to develop in them a high degree of reliability insofar as their predictive value is concerned. If new tests have to be designed each year, their quality will decrease, their cost will increase, and the times at which they can be administered will decrease.

The Graduate Record Examination is important in the selection of graduate students. It is the only common denominator we have at our disposal. Grades from one university are not comparable to those from another nor does a letter of recommendation from one person have the same meaning as that from another. My eight years of experience as dean of a graduate school have shown that when graduate students run into difficulty in their comprehensive examinations or have difficulty in producing a satisfactory dissertation, their performance on the Graduate Record Examination has been marginal.

It is an inordinate waste of time and money, both on the part of the student and the university and its sources of support, to admit students to a graduate program and then dismiss them for academic failure a year or two later. I am convinced that the results of the Graduate Record Examination help us to keep such mistakes at a minimum.

You may share this letter with other members of the House of Education and Labor Committee. Thank you for your attention.

Sincerely yours,

C. G. Stuckwisch
Dean for Graduate Studies
and Research

CGS/g
cc: Dr. Michael J. Pelczar
Council of Graduate Schools
Dr. James J. Hudson, President
Council of Southern Graduate Schools

A private, independent, international university
An equal opportunity/affirmative action employer

WASHINGTON STATE UNIVERSITY
PULLMAN, WASHINGTON 99164

THE GRADUATE SCHOOL

SEP 24 1979

September 19, 1979

Michael J. Pelczar, Jr.
President
Council of Graduate Schools in the U.S.
One Dupont Circle
Washington, D.C. 20036

Dear Dr. Pelczar:

Thank you for the August 17 package outlining pending legislation aimed at standardized tests. I share your concerns about the legislation: what is, on the surface, an admirable attempt to force testing agencies to inform "consumers" of the steps in an important academic process will in fact destroy the common use of standardized tests. For many of our departments, ETS examinations are important elements in the admission process. Although all of us recognize that standardized exams are not perfect, they certainly are helpful indicators of an applicant's potential when considered with other criteria. Losing the use of standardized tests will make admissions decisions more difficult.

Certainly we *shall* lose the use of the exams to a great extent. As several of the legislation's opponents have pointed out, the cost of the exams will undoubtedly rise dramatically and the quality of the exams will drop. Because of the higher cost of the exam, fewer students will take it. Many departments will no doubt react to the higher cost by no longer requiring the exam. Because the validity of the exam will be seriously compromised by continual *total* revision, few departments will wish to use it.

We are not naive. I repeat, we *know* that the present set of standardized exams are not perfect; consequently, we do not use results from those exams exclusively for admissions decisions. But the exams are a useful tool when used in conjunction with other useful items. Just as a carpenter is not rendered completely ineffective if one tool is removed from his box, so will our admissions committees not collapse completely without standardized tests; however, like the carpenter, our committees will not work as effectively as they would were they allowed to use all of their tools.

Please record us, then, as being opposed to the Weiss and Gibbons bills.

Sincerely,

Richard D Fulton

Richard D. Fulton
Assistant to the Dean

RDF/jj

cc. Dr. White
Dr. Khang.

THE UNIVERSITY OF ALABAMA
THE GRADUATE SCHOOL
POST OFFICE BOX W
UNIVERSITY, ALABAMA 35486

OFFICE OF THE DEAN

August 31, 1979

The Honorable John Buchanan
United States House of Representatives
2159 Rayburn House Office Building
Washington, D. C. 20515

Dear Mr. Buchanan:

I am taking this opportunity to express to you my opinion on proposed legislation aimed at standard education tests, the Weiss bill (HR4949) and the Gibbons bill (HR3564) the so-called "truth-in-testing" bills which are being considered by the House Education and Labor Committee.

Although the provisions of these bills might seem noncontroversial, they could have drastic effects upon students and graduate programs. Principally these would be:

1. Costs to the test agencies and eventually to students would rise dramatically since a completely new test would have to be developed for each examination date and the existing method of pretesting or validating future questions would have to be replaced with a more expensive system. I am sure that the mechanisms of these circumstances have been brought to your attention.

2. Students will have fewer opportunities to take the examinations since increased difficulty in constructing the tests and the impossibility of reusing questions will greatly restrict the number of times during the year that the examination can be administered. For example, to accommodate student demand, the Graduate Record Examination is offered six times per year. A reduction in number of times offered to two - a meaningful probability - would significantly reduce student freedom and impair a nationally established graduate school admission process.

3. Quality of the examinations would deteriorate. There is a limit to the number of good questions which can be developed and validated, and as the quality of the test questions decline the value of the examination as a component of the graduate admissions process will be limited.

4. The Test of English as a Foreign Language, a "standardized test," is used widely to determine the ability of foreign students to utilize the English language in American universities. At a time when United States higher education is being called upon, increasingly, to educate students of other countries, a compromise of the effectiveness or validity of this examination could result in our universities admitting students without truely adequate English competency, thus compromizing the teaching and learning environment for the foreign students and our own English speaking students.

5. With the wide variation in academic programs and grading systems existing throughout the country, it is essential that some form of frequently administered, well validated national examination be available for use in the graduate admissions process. Like other schools, The University of Alabama does not rely on national examinations alone, but they are an important item in the total group of factots which are considered in determining which students have the potential to succeed in graduate studies and are admitted thereto. Unfortunately, no suitable substitute exists at the present time and a reduction in frequency and quality of these well established examinations would impair our ability to assess a student's graduate potential.

I hope my comments may be of some value to you as you consider HR3564 and HR4949, and I would be happy to provide any additional information you might require on this subject.

Sincerely yours,

William H. Macmillan
Dean

WHM:wm

bcc: President David Mathews
Dr. Richard Thigpen
Colonel Floyd H. Mann
Dr. Michael J. Pelczar, Jr.

cc White
Kenney ✓

MSU
OFFICE OF THE DEAN
COLLEGE OF GRADUATE STUDIES MONTANA STATE UNIVERSITY, BOZEMAN 59717

September 13, 1979

Congressmen Pat Williams, Montana
House Education and Labor Committee
U.S. House of Representatives
Washington, D.C. 20515

Dear Congressman Williams:

Having perused two bills pending before Congress, namely the Weiss Bill (H.R. 4949) and the Gibbons Bill (H.R. 3564), I should like to voice my negative concern about both of them.

We here feel that any federal oversight of standardized testing poses grave dangers to the entire integrity of the system. This kind of centralization and regimentation is bound to limit the frequency of testing and to compromise the quality of the examinations, since the Weiss Bill calls for disclosure of all test forms.

In short, while admirable in their attempts to assure fairness through federal regulations, the bills seem sure to jeopardize the reliability and availability of the examinations themselves.

Thanks for your consideration.

Yours Vary Sincerely,

Michael P. Malone
Dean of Graduate Studies

MPM/me

cc: Michael J. Pelczar

TELEPHONE (406) 994-4145

cc White
Kraus

ST. CLOUD STATE UNIVERSITY
SCHOOL OF GRADUATE STUDIES

Graduate Studies (612) 255-2113
International Studies (612) 255-4287
St. Cloud, Minnesota 56301

August 24, 1979

Mr. Michael J. Pelczar, Jr.
President
Council of Graduate Schools
One Dupont Circle
Washington, D. C. 20036

Dear Mr. Pelczar:

My name is Lowell Gillett. I am Dean of the Graduate School at St. Cloud State University. I am writing to express my concerns relative to the so-called "Truth in Testing" measures that are currently under consideration by the Congress (H.R. 4949 and H.R. 3564). As you know, St. Cloud State University is a state supported institution. At the present time, we do not restrict admission to our graduate school programs to any specified number of students. We accept all who qualify, with this decision based on several factors: past academic performance, potential performance (for which we use the Graduate Record Examination), or proven performance in a professional capacity in which we seek to obtain evidence that the student's combination of ability and motivation are indeed likely to result in successful graduate study. Graduate education is expensive and we feel that admitting to graduate school only those students who can provide evidence of potential success is our obligation to the taxpaying public. Our philosophy is to provide graduate opportunities to all who can demonstrate their ability to profit from advanced study. But we believe that we must provide every possible option to the student to demonstrate that potential.

The "Truth in Testing" legislation, if enacted, would undoubtedly render standardized tests such as the Graduate Record Examination (GRE) relatively useless as a measure of potential, since coaching for a test, using actual test questions, will render the results worthless. And we will no longer utilize this option if we consider the information to be worthless. On the other hand, the present forms of coaching (which appear to have generated many of the pressures for such legislation), appear to be largely focused upon increasing the student's facility in the use of the English language and upon improving his/her mathematical skills and understanding. The higher scores resulting from this kind of coaching are based on increases in the student's knowledge and understanding of the subjects which are the foundation of all formal learning. Therefore, these are efforts that, should they increase test scores sufficiently to result in the admission of students who would otherwise be denied, are nevertheless truly beneficial to the student's actual success in graduate study. The abilities which are improved are, in fact, among those which contribute to successful

SCSU is an equal opportunity / affirmative action employer and is in compliance with all state and federal [illegible]

academic work. In many respects, then, the present coaching efforts are actually beneficial, not only to the student through admission to graduate school, but to the society.

On the other hand, if the concern is that the present coaching activities benefit only those whose financial means permit their participation, surely any legislation which seems certain to increase the use of costly "coaching" programs will further reduce the access to higher education of lower income persons. Unfortunately, a high proportion of these persons are members of minority groups. Further, with test questions available, the coaching will certainly stress the test questions themselves, rather than attempting to raise the foundation knowledges and skills which are necessary for academic success.

In my twelve years as Graduate Dean, I could cite numerous examples of students who would be seriously harmed by the elimination of this valuable tool, that is, they would not have been accepted into graduate school. But my personal case is an excellent example. I was an immature late-bloomer. My high school record was modest. I was admitted to Gustavus based solely on my standardized test scores. In college, I had still not settled down and my academic performance was only slightly better than mediocre. I eventually applied for admission to graduate school at the University of Minnesotra and was denied. I persevered and was ultimately accepted provisionally, again based on my performance on standardized tests. Without this option being available, I would not have been given the opportunity to pursue graduate study.

If we are to provide people the opportunity to maximize their productive potential, we need all the evidence possible to assess the potential of each applicant. The individual should be given every opportunity to demonstrate his/her potential for success, and standardized tests provide a useful tool in that assessment.

I urge your active opposition to these bills--bills which will seriously reduce or eliminate the possibility of utilizing standardized tests productively. As with so many things which have occurred in modern society, I can confidently predict that in this case, the proposed cure will do more harm than the disease.

Yours very truly,

Lowell R. Gillett

Lowell R. Gillett
Dean, School of Graduate Studies

aes

Illinois State University
President

cc [illegible]

September 6, 1979

The Honorable Carl D. Perkins, Chairman
House Education and Labor Committee
U. S. House of Representatives
Washington, D. C. 20515

Dear Representative Perkins:

My purpose in writing is to oppose the proposed legislation of The Weiss bill (H.R. 4949) and The Gibbons bill (H.R. 3564), the so-called "Truth in Testing" legislation and similar to The New York State bill (S. 5200, LaValle et al.) which was recently enacted.

In opposing this proposed legislation I am not unaware that some misuse of standardized testing does occur. The solution to the problem of improper use of standardized tests, however, is not to enact legislation as proposed in the Weiss and Gibbons bills. The disadvantages of these two bills far outweigh the advantages. The disadvantages are several, and each is substantive as a potential disservice to graduate education in the United States. First, and primary is the threat to test quality. It may not be possible to maintain the present quality and value of the test as an aid for decision-making for student admissions. This is particularly true for the immediate future when the requirement for disclosure of test questions and answers will result in compromised tests.

Second, the proposed legislation, if enacted, would increase student costs. The constant development of new tests, which is extremely difficult, is expensive. Third, the proposed legislation would probably result in fewer test administrations, thereby decreasing access to taking the test for the student. This could adversely affect admissions processes.

Your consideration of these disadvantages in the hearings on October 10 and 24 will be appreciated. In my opinion, the legislation should not be enacted.

Sincerely,

Lloyd I. Watkins
President

LIW/gc

Normal-Bloomington, Illinois
Phone: 309/438-5677

Hovey 308
Normal, Illinois 61761

Equal Opportunity/Affirmative Action University

AUBURN UNIVERSITY

AUBURN ALABAMA
36830

THE GRADUATE SCHOOL

Office of the Dean

Telephone 826-4700
Area Code 205

September 14, 1979

The Honorable Bill Nichols
2417 Rayburn Building
Washington, DC 20515

Dear Bill:

I am deeply concerned about two Bills presently under consideration by the House Committee on Education and Labor. These are: The Educational Testing Act of 1979 (H.R. 4949) and the Truth in Testing Act of 1979 (H.R. 3564). Some of the information requested in these Bills will certainly be helpful to applicants required to submit scores on standardized tests for admission to undergraduate and graduate programs in institutions of higher education. In fact, it is my impression that much of this information is already available. Some of the provisions constitute what I believe to be information overkill. That is to say the information might be interesting, but it is not critical to the objectives of the legislation, and it would impose an additional, unnecessary burden on those affected by the legislation. And parts of the proposed legislation will cause universities great difficulties with the already difficult task of making admission decisions. This is especially true under Section 5 of House Bill 4949. This legislation will require that agencies offering standardized tests release copies of all test questions and answers to the public within 30 days after the results of the test are released.

Let me make several points to underscore the difficulties caused by this section of the proposed Bill. Universities are faced with two alternatives in making admission decisions. They can adopt an open admission policy and admit all persons that apply to graduate and undergraduate programs, or they can attempt to limit admission to those students that appear to have the best aptitude and chance of success in programs of higher education. Most universities in this country have opted for the second approach because limited resources require that admission to most of their programs be restricted in some way and because the quality of graduate and undergraduate programs is directly dependent upon the academic ability of the students enrolled in those programs.

A LAND-GRANT UNIVERSITY

The question of criteria to be used in making admission decisions is in itself a very difficult problem. Basically universities depend on grade-point averages, scores on standardized tests, and such additional items as letters of recommendation and interviews. As you are well aware, the value of a grade-point average as a criterion for admission is directly related to the quality of the school where the grade-point average is earned. It is impossible to standardize grade-point averages, and I hardly need tell you about the subjective nature of letters of recommendation and interviews. So the only tool available to university admission officers that offers some objective basis of comparison between the various candidates applying to the university are standardized tests such as the Graduate Record Examinations (GRE) used by many graduate schools.

The GRE is administered to thousands of students each year and as I noted above, offers an objective basis on which to compare a student's aptitude for graduate study on a national scale. Furthermore, the aptitude tests of the GRE measure such skills as reading comprehension, basic mathematics, data interpretation, logical analysis, and ability to comprehend relationships. Furthermore, these aptitude tests have been shown through a number of studies to be reasonable predictors of success in academic programs. It should also be pointed out that most universities use these tests in conjunction with other information such as graduate grade-point averages and letters of recommendation when making admission decisions. At Auburn we have found that there is a reasonably good correlation between success in graduate school as measured by graduate grade-point averages and a weighted admission score derived from GRE scores and undergraduate grade-point averages. I don't want to belabor this point, but I do want you to understand the importance of standardized tests in making admission decisions in universities.

In order to standardize these tests, they are given to large numbers of students in different locations and the questions used are carefully constructed, evaluated, and pre-tested before inclusion in the regular examination. This is a time-consuming and expensive process. If, following each administration of the examination, it becomes necessary to release all the questions used and the answers, then the testing agencies will have to prepare completely new questions for each examination. This will have several effects. It will greatly increase the cost of the examinations to prospective students. It will decrease the precision and quality of the questions included in the examination. It will likely lead to fewer administrations of the

The Honorable Bill Nichols
September 14, 1979
Page 3

examinations--greatly inconveniencing many applicants for graduate and professional programs. And, it will most surely decrease the reliability and validity of the standardized tests.

I ask your careful consideration of these factors as you determine your position on these two pieces of legislation. If I can be of assistance to you in your deliberations, please call on me. Thank you for your continued support of higher education in the State of Alabama.

Sincerely,

Paul F. Parks
Dean

PFP:mm

SEP 17 1979

HOWARD UNIVERSITY
WASHINGTON, D. C. 20059

THE GRADUATE SCHOOL OF ARTS AND SCIENCES
OFFICE OF THE DEAN

September 11, 1979

Dr. Michael J. Pelczar, Jr.
President
Council of Graduate Schools in the U.S.
One Dupont Circle
Washington, D.C. 20036

Dear Dr. Pelczar:

Your communication concerning Truth-in-Testing legislation was received. The Graduate School of Arts and Sciences is concerned about the proposed bills even though the results of standardized tests are used at Howard University primarily for diagnostic and research purposes rather than for admissions.

Our analysis of the proposed legislation and related material can be summarized in the following manner:

ADVANTAGES OF TRUTH-IN-TESTING LEGISLATION

1. Passage of such laws might be helpful in enlightening students and parents and reducing present ignorance about such tests.
2. Such legislation might cause publishers and administrators of these tests to be more circumspect in their construction and interpretations of standardized tests.
3. Educational institutions might move in the direction of increasing the weight given to other factors in their admissions processes.

DISADVANTAGES OF TRUTH-IN-TESTING LEGISLATION

1. The cost for students will increase substantially.
2. The number of students taking the test and the number of institutions requiring it may decrease considerably.
3. The new demands placed on testing agencies are so great that the number of annual testing administrations may be markedly reduced. The quality of the tests may be affected adversely.

DISADVANTAGES (continued)

4. Much confusion, animosity and legal action may result because of different interpretations of correct answers, marking procedures, and applicability of results.
5. Most of the proposed laws have not provided sufficient time before voting for public consideration and reaction.
6. Passage of these laws might result in fewer disadvantaged students taking the test.

As the above indicate, our current belief is that the proposed Truth-in-Testing legislation contains more disadvantages than advantages to higher and professional education.

Sincerely yours,

Edward W. Hawthorne, M.D., Ph.D.
Dean

EWH/jeh

THE UNIVERSITY OF NORTH CAROLINA
AT GREENSBORO

OCT 1 1979

School of Health, Physical Education
and Recreation
Office of the Dean

September 26, 1979

The Honorable Ike F. Andrews
U. S. House of Representatives
Washington, D. C. 20515

Sir:

I am writing to express my concern over the pending legislation aimed at standard tests. It is, without question, true that the legislation would probably cause a deterioration of test quality, increased costs, and less convenience to the persons involved. My greater concern, however, relates to what this legislation would do to the validity of the tests. No test is perfect, but decisions made on the results of good tests are wiser than sheer guesses. I hope you will exert your influence to turn the tide of this legislation which would hamper educational decision making at all levels.

Sincerely,

Rosemary McGee
Rosemary McGee
Acting Dean

RM:ll

bc: Michael J. Pelczar, Jr.

GREENSBORO, NORTH CAROLINA/27412
THE UNIVERSITY OF NORTH CAROLINA *is composed of the sixteen public senior institutions in North Carolina*
an equal opportunity employer

SEP 26 1979

The University of Vermont
OFFICE OF THE DEAN, GRADUATE COLLEGE
WATERMAN BUILDING, BURLINGTON, VERMONT 05401
(802) 656-3160

September 24, 1979

The Honorable James M. Jeffords
U.S. House of Representatives
Washington, D.C. 20515

Dear Mr. Jeffords:

I am writing to express my reservations about the proposed bill H.R. 4949, The Weiss bill and companion proposals, to require certain information be provided to individuals who take standardized educational admissions tests. My reservations are based on the following observations:

.Such legislation as proposed above will increase the costs, primarily to those taking the tests, and more importantly may exhaust the various ways in which a specific set of items related to an important concept or skill can be expressed. For example, questions dealing with spatial relations, sequential reasoning, and transferability of information from one specific situation to another can be framed in a finite and small number of questions without giving away the potential discriminability of the item.

.Results from standardized tests make up only one component of the admission's file of an applicant to graduate or professional schools. The extent to which academic grades and standard test scores are congruent provide degrees of confidence in an admissions decision that would be compromised if such test items became available to potential applicants. I suspect the relationship between grades and test scores will become increasingly important to decisions on admissions as the interval between completion of undergraduate education and application to graduate or professional schools increases as has been the case in recent years.

.There are a number of tutoring techniques and procedures available in preparing for standardized admissions tests which appear to influence (albeit modestly) test performance.

An Equal Opportunity Employer

> Despite such modest changes of test performances, one of the objectives of standardized admissions tests is to assess potentials and aptitudes fashioned from the previous, and cumulative educational experiences extending over years of prior education. Accordingly, the availability of test items may indeed compromise an important index of such aptitudes and potentials.

I ask you not to support the above legislation or comparable proposals. I will be glad to answer any questions about the above upon your request (802-656-3160).

My best wishes to you.

Sincerely yours,

Robert B. Lawson
Associate Vice President for Research
and Dean of the Graduate College

RBL:lbm

cc: President Lattie Coor
University of Vermont
✓ President M. Pelzar, CGS

CLEMSON
UNIVERSITY

OFFICE OF GRADUATE STUDIES AND UNIVERSITY RESEARCH

OCT 5 1979

October 3, 1979

Congressman Butler Derrick
United States House of Representatives
Washington, D.C. 20515

Dear Butler:

I certainly enjoyed your remarks to the Sertoma Club yesterday. Everyone appreciated the candor with which you addressed each issue, including your pay raise.

The "Truth in Testing" legislation presently proposed by two bills, HR 3564 and HR 4949 is of great concern to the higher education community. I mentioned to Mike that I would send you several items from my file on this subject for your review. In ten words or less, it is bad legislation and will have a devastating effect on undergraduate and graduate admissions in colleges everywhere. The experience that New York State is having should be examined carefully before such chaos is allowed to spread across the entire nation.

Enclosed is information on the New York State legislation and some of the reaction to it. Also attached are the two House bills along with testimony given before the subcommittee and a letter from Lyle Jones, Vice Chancellor of the University of North Carolina.

I hope you will agree that the legislation is ill-conceived and is not in the best interest of quality higher education. Any action you can take to defeat these bills would be a worthy service to the educational community.

Sincerely yours,

Arnie

A. E. Schwartz
Dean of Graduate Studies
and University Research

AES:dm
Attachments

ccs: President Bill L. Atchley
Dr. Michael J. Pelczar, Jr.
President, Council of Graduate Schools in the U.S.

CLEMSON, SOUTH CAROLINA 29631 • TELEPHONE 803/656-3195

Boston College, Chestnut Hill, Massachusetts 02167 Telephone (617) 969-0100

Graduate School of Arts and Sciences
Office of the Dean

October 22, 1979

The Honorable Carl D. Perkins
Committee on Education and Labor
Subcommittee on Elementary, Secondary &
Vocational Education
B-346C Rayburn House Office Building
Washington, D.C. 20515

Dear Mr. Perkins:

In response to your kind letter of October 9th indicating that you would be pleased to accept a written statement from me for inclusion in the hearing record in opposition to the two bills dealing with standardized testing, H.R. 3564 and H.R. 4949, currently before your Subcommittee, may I associate myself with the position taken by the Association of Graduate Schools in American Universities and the letters of countless graduate deans to your Subcommittee.

All of these without exception urge that action upon these bills at a very minimum be deferred until we may have the benefit of the experience under the recent testing legislation adopted by New York State.

As presently drawn, these bills are both premature and defective and should not be passed.

Sincerely yours,

Donald J. White
Donald J. White
Dean

sc

NATIONAL
1937

national association of college admissions counselors
9933 Lawler Avenue, Suite 500, Skokie, Illinois 60077, Tel. 312/676-0500

[illegible] EXECUTIVE BOARD

[illegible]

STAFF

[illegible]

October 23, 1979

The Honorable Carl D. Perkins
The House of Representatives
2365 Rayburn House Office Building
Washington, D.C. 20515

Dear Mr. Perkins:

We are writing on behalf of the National Association of College Admissions Counselors to express our deep concern with legislative proposals, H.R. 3564 and H.R. 4949. The subject of these two proposals was debated by 1300 members at our recently concluded National Conference. We are especially concerned with how these bills will affect the student or the education consumer.

It is with the following purpose in mind that we ask your committee to withhold a decision on these bills until the next session of Congress for the reasons listed below.

NACAC is a not-for-profit association of 3,000 high school counselors and college admissions officers. One of our articles of incorporation is "to serve students, parents, secondary schools, postsecondary educational institutions and related educational organizations and agencies by considering the whole range of influences on a student's transition from secondary school to further education without regard to race, creed, sex, political affiliation or national origin."

Specifically:

1. While NACAC fully endorses students' rights to free and full information, we are concerned that such access not result in either limiting a student's educational opportunities or dramatically increasing the costs of tests to students.

2. The impact of state laws on the testing issue (especially New York State) has not had an opportunity to be fully "tested". There should be sufficient lead time to prepare for changes before national laws are imposed upon the admissions process. Consideration must be given to the ramifications these bills will have, if enacted.

In July, two major testing agencies stopped administering tests in New York State and it now appears that twenty more test sponsors will follow suit. The College Board recently revised its 1980 schedule for administering the SAT in the same state. The result: fewer tests and higher costs to the student. Further, New York State has moved up the date for implementing its new law to July 1, 1980 instead of January 1 as previously scheduled. Reason for the action is to give the New York State Assembly an opportunity to again review and revise, if necessary, the approved legislation.

3. The secondary school counselor will bear the burden of any changes in the law as will the student. Federal regulation is not needed at this time on a system that is already overburdened with the new laws and procedures on financial aid and the Privacy Act.

4. If enacted as written, the new law would require more counseling on behalf of the student at a time when most local school budgets are being cut back in the area of student counseling services.

5. The logical approach to correcting test abuse is through education rather than through legislation. Training all counselors and admissions officers who use tests each day to help or make decisions on behalf of students is a national necessity. Professional associations like NACAC see this education process as a form of self regulation.

6. Our society needs to deemphasize the use of tests in the admissions decision process. The proposed laws tend to highlight testing and take it out of perspective in the admissions process.

7. If enacted, the law will increase the opportunity for nonprofessionals to profit from questionable services regarding test taking.

We have stated these seven reasons because we do not believe your committee has received any testimony from the professionals who work with students and parents on a daily basis.

Before any bill is sent to the full committee, NACAC, representing the professional counselors at both the secondary and postsecondary levels, stands ready to work with you and your colleagues on how new testing laws can be written to enhance the educational value of the tests and enable students to have free and full information on tests.

Enclosed are three NACAC documents for your information - Statement of Students' Rights and Responsibilities in the College Admissions Process, Statement of Principles of Good Practice, and the Monitoring Procedures for these statements.

Sincerely yours,

Delwin Gustafson

Delwin K. Gustafson
President

Charles A. Marshall

Charles A. Marshall
Executive Director

CAM:mg
Enc.

cc: Members of Subcommittee on Elementary, Secondary and Vocational Education

Representatives Weiss, Miller and Chisholm

MAJORITY MEMBERS:
WILLIAM D. FORD, MICH., CHAIRMAN
JOHN BRADEMAS, IND.
FRANK THOMPSON, JR., N.J.
MARIO BIAGGI, N.Y.
PAUL SIMON, ILL.
PETER A. PEYSER, N.Y.
JOSEPH M. GAYDOS, PA.
AUSTIN J. MURPHY, PA.
TED WEISS, N.Y.
WILLIAM R. RATCHFORD, CONN.
DON BAILEY, PA.
CARL D. PERKINS, KY., EX OFFICIO
[illegible]

CONGRESS OF THE UNITED STATES
HOUSE OF REPRESENTATIVES
COMMITTEE ON EDUCATION AND LABOR
SUBCOMMITTEE ON POSTSECONDARY EDUCATION
318 HOUSE OFFICE BUILDING ANNEX #1
WASHINGTON, D.C. 20515

MINORITY MEMBERS:
JOHN H. BUCHANAN, JR., ALA.
JAMES M. JEFFORDS, VT.
MICKEY EDWARDS, OKLA.
THOMAS J. TAUKE, IOWA
THOMAS E. PETRI, WIS.
JOHN M. ASHBROOK, OHIO, EX OFFICIO

October 24, 1979

Honorable Carl D. Perkins
Chairman
Subcommittee on Elementary, Secondary
and Vocational Education
B-346-C Rayburn HOB
Washington, D.C. 20515

Dear Mr. Chairman:

As the ranking Member of the Subcommittee on Elementary, Secondary and Vocational Education, and the Chairman of the Subcommittee on Postsecondary Education, I have received considerable correspondence on H.R. 3564 and H.R. 4949, which address the issue of educational testing.

Of the correspondence I have received, the attached letter from Dr. Ira Polley, Assistant Provost for Admissions and Records at Michigan State University, was particularly enlightening. Dr. Polley makes several excellent points, which I would like to share with you and have included in the hearing record for these two bills.

I am sure you will also find his observations most helpful in your analysis of the legislation.

With kind regards, I am

Sincerely,

William D. Ford

WILLIAM D. FORD
Chairman

WDF:rsn

Enclosure

cc: Dr. Ira Polley

MICHIGAN STATE UNIVERSITY

OFFICE OF THE PROVOST
EAST LANSING · MICHIGAN · 48824

October 12, 1979

The Honorable William D. Ford
House of Representatives
2368 Rayburn House Office Building
Washington, D.C. 20515

Dear Bill:

As a member of the Subcommittee on Elementary, Secondary and Vocational Education, House Committee on Education and Labor, you may be interested in the reactions that I have to H.R. 3564 (the Gibbons' bill) and H.R. 4949 (the bill introduced by Mr. Weiss for himself, Mrs. Chisholm and Mr. Miller of California).

I can understand some of the concerns that have led a number of fine people to believe there are grave problems with admissions tests that can be corrected only with national testing legislation. But in my many years in educational posts of one kind of another, I really have not seen convincing evidence that there are abuses that victimize students and parents. It may be that a few higher educational institutions place too much reliance on some tests, but very few universities or colleges make decisions on admissions by according great weight to national tests. We at Michigan State University do use the SAT and the ACT tests, but neither plays the decisive role in our work. High school grades in academic subjects, the number of "solid" subjects taken, and the trend in the grades are the decisive factors. If a few colleges do place inordinate reliance on national tests, parents and students have a remedy: That is, since there is such a large number of excellent public and private universities and colleges, no one has "to do business" with any college that may believe tests give infallible guides. As a matter of fact, I personally know of no university or college that operates on the assumption that test results are the decisive element in the decision-making process.

The Gibbons' bill, as you know, provides, in part, that everyone who takes the test is to be informed of "the score which is generally required for admission to institutions of higher education". There is here an assumption that is based on no facts at all. As my earlier comments have suggested, the system of higher education in America is a pluralistic one. There is no test score that all institutions, fifty percent of the institutions, or even three percent of the institutions use. Indeed a single institution looks at test scores along with many other factors, and accordingly has established no score that guarantees admission or that results in a denial.

The Honorable William D. Ford
Page 2
October 12, 1979

Let me say a few words on one feature of the bill introduced by Mr. Weiss and others. It contemplates that any one who takes a test has the right to a copy of the test and the correct answers. This provision is filled with mischief. Some of the clear consequences are as follows. *First*, its operation would make difficult the improvement of tests. I recall from my days, years ago, as a teaching assistant in Political Science at the University of Minnesota. In the large American government courses, with enrollments of a few thousand, multiple choice questions were used. We had assistance from a testing office in the University to do validity studies, so that imperfect questions including those that do not discriminate could be eliminated. For example, if 95 percent of the students know that Congress is a bicameral body, that question isn't of much value (it does not discriminate). Or if only five percent of the students can identify 1930 as the date of the Norris-LaGuardia Act, again that question may not be suitable. To return to the Weiss bill, I do not see how validation studies and test improvements could be prosecuted, if all tests and right answers are to be published. And this leads me to my *second* point. The tests and answers would indeed become published by tutoring organizations that would be giving assurances that they can guarantee your youngster will perform magnificently on a test. In this connection, please see the excellent letter (copy enclosed) that Dr. Kenneth B. Clark, the author of *Dark Ghetto*, wrote to the *New York Times*.

I have a *third* observation about the Weiss bill. If it is passed, the people who will suffer are the parents and the students who will have to pay substantially more for national tests.

To summarize, it is my profound conviction that there are no convincing reasons why Congress should enact national testing legislation. As a matter of general philosophy, I share many of the views that Mr. Nader has. But I do not believe that either the Gibbons or the Weiss bill should be supported on the mistaken idea that it is an enlightened example of a measure to protect consumers. It most certainly will not benefit parents and students who seek to enter undergraduate or graduate programs.

If my comments provoke questions in your mind, I will be happy to respond. Or if your staff colleagues wish to pursue any of the points I have tried to make, I'll be happy to assist them.

All best wishes.

Sincerely,

Ira

Ira Polley
Assistant Provost for Admissions and Records

IP/jh

Enclosure

[From the New York Times, August 18, 1979]

'This So-Called Truth-in-Testing Law Is a Placebo'

To the Editor:

On July 24, The Times uncritically endorsed the "truth-in-testing" legislation, which was signed by Governor Carey. Your editorial stated that this amendment to the education law would take "the mystery out of college testing" and that "students deserve to know how they are being rated and judged."

In spite of the good intentions of your editorial, it was misleading. So was Ralph Nader's assertion in an Aug. 3 letter, that the legislation will require test companies "to explain to students what the scores mean and how they will be reported to schools." And Nader's claim that objections to this legislation reflect only the "corporate style lobbying efforts of testing services" is inaccurate.

The Board of Regents and the Commissioner of Education of the State of New York independently advised the Governor against signing this legislation. It is my contention that this so-called truth-in-testing law is a placebo.

However laudable its intentions, this law cannot force test companies to explain the meaning of test scores to students; and certainly this law cannot deal with the complex issues of test validity and the role of cultural factors in influencing test results. The construction, evaluation and interpretation of tests are highly technical matters which must be dealt with by ongoing research by those who are trained in this specialty. The important problem of the use and abuse of standardized tests cannot be resolved by a simplistic law which confuses this issue with consumer-protection problems.

Admission tests measure the abilities developed by individuals over a long period of time, both in and out of school. The value of test results is directly related to the equality of opportunity afforded to the candidates taking a particular test. The assumption that a candidate's examination of a test already taken will somehow improve scores or reduce the chances of abuse is highly questionable. One can further assume that only the more privileged students will avail themselves of the opportunity to examine the test materials. These students will seek private tutoring on the assumption that this will increase their future test scores. Those individuals who cannot afford this privilege will certainly not improve their test scores by merely examining the test questions and scores. This is a deceptive and meaningless exercise.

The New York State version of the truth-in-testing law is misleading and confuses the public. It will not benefit minority and poor students. It could, in fact, be detrimental to already-disadvantaged individuals. Appropriate legislation in the general area of the abuse and the premature exclusionary use of standardized tests might be desirable, but such legislation should be the culmination of serious inquiry rather than a political gesture.

It is my firm belief that this is such an important matter that the New York State Legislature should hold a series of hearings with concerned professionals for the purpose of amending this law, which promises much more than it can possibly deliver.

KENNETH B. CLARK
Member, Board of Regents
New York, Aug. 7, 1979

October 24, 1979

The Honorable Carl E. Perkins
Chairman, Elementary, Secondary and
Vocational Education Subcommittee,
House Education and Labor Committee
B-346 C Labor and House Office Building
Washington, DC 20515

Dear Mr. Perkins:

As per your letter of October 9, 1979, I am writing several statements in response to two bills dealing with standardized testing. H. R. 3564 and 4949. My opinions are my own and have been formulated by my working with standardized testing programs for six years. My position is Coordinator of Testing at Miami-Dade Community College, North Campus. I administer one of the largest testing programs (over 50 standardized tests) in the United States for the national testing agencies (Educational Testing Service, American College Testing program and Psychological Corporation).

I will summarize my opinions in a series of statements:

1. I feel that the passage of the bills will not be in the best interest of the population they intend to serve.
2. The cost of the admission tests will be increased greatly, thus reducing access to colleges and professional programs by several categories of students. The testing agencies will simply reduce the number of test sessions, thus causing the test centers to increase the capacity of testing sessions. Many schools and universities will not be able to accommodate this increase.
3. As a professor teaching graduate courses in tests and measurement, I cannot see the benefits that would be served by sending original answer sheets and copies of test items to students. Research studies have shown some fluctuation in scores, but within certain limits, I think the tests are valid predictors of success in college and professional programs.

-2-
Mr. Carl E. Perkins
October 24, 1979

4. If monitoring of standardized testing programs is necessary, I propose that this be done by the newly created Department of Education. The testing agencies abide by certain constraints with respect to items and item pools. Professional admission councils monitor the programs now. Research indicates that there is a high correlation, generally, between test scores and success in colleges and graduate studies.

5. Finally, I wonder if the testing agencies can, within the time frame specified, produce entirely different forms of standardized admission tests, equate and norm them for comparability of scores. The fact that students have access to the actual test items will change certain factors in the process of standardized testing.

6. Before these bills are acted upon, I think some research studies need to be done in New York State, showing the implications of this type of legislation.

Thank you for your time in this matter.

Sincerely,

John R. Scerba

John R. Scerba, Ph. D.
11822 N. E. 12th Court
North Miami, Florida 33161
(305) 685-4412

JRS/ec
cc: Congressman Gibbons
Congressman Weiss

Attachments

AMERICAN
MEDICAL
RECORD
ASSOCIATION

875 NORTH MICHIGAN AVENUE • SUITE 1850, JOHN HANCOCK CENTER, CHICAGO, ILLINOIS 60611 | (312) 787-2672

EXECUTIVE OFFICE

October 30, 1979

Honorable Carl D. Perkins, Chairman
Committee on Education and Labor
U.S. House of Representatives
2181 RHOB
Washington, D.C. 20515

Dear Mr. Perkins:

In behalf of the 23,000 members of the American Medical Record Association, (AMRA), I take this opportunity to comment on the pending proposed federal test legislation - H.R. 3564 and H.R. 4949. I understand that hearings were held recently on this legislation and I wish to enter these remarks as part of the Committee hearing proceedings.

First, AMRA supports the institution of standardized testing and the concept of fair reporting practices to consumers. We are in support of requirements of H.R. 3564 and H.R. 4949 to require: description of the test purpose, description of subjects to be tested, margin of error and extent of reliability, distribution of test scores, applicants rights to obtain test scores and details of such disclosure. These procedures are sound practices which the American Medical Record Association currently follows in its National certification testing program.

Secondly, AMRA recognizes and supports the private sector nationally endorsed standards, safeguards, and procedures established for standardized test development and administration. Such standards have been set by professional testing and measurement specialists and have resulted in reliable test instruments. We favor maintaining private sector control over testing development and procedures and oppose the submission of test questions, scores, and scoring procedures to the U.S. Commissioner of Education. We oppose entry of the federal government into these procedures, as this will lead to government control of educational testing in the United States.

Third, AMRA opposes any legislation that would mandate disclosure of questions and answers of standardized tests to the public or test applicant. We oppose release of this information because:

1) The value of standardized test questions used across time will be lost.
2) The "pool" of test questions, especially for specialized academic areas, would diminish over time; and, the availability of the question pool in the public arena would provide unfair advantage to applicants with ready access to the "pool". Thus, the high reliability of standardized tests to test knowledge or ability would diminish.
3) The cost of testing will increase due to the necessity to prepare completely new tests each time a test is given -- in many instances several times a year.

In summary, AMRA supports fair testing development and administration procedures which have been developed by the private sector and which are supported in H.R. 3564 and H.R. 4949.

We stand opposed, however, to the submission of test questions and answers to the federal government and test applicants.

Thank you for the opportunity to submit these comments for the record.

Sincerely,

Carolyn Cave

Carolyn Cave, Ph.D., RRA
Executive Director
AMERICAN MEDICAL RECORD ASSOCIATION

CC:JR:lm

ED BETHUNE
2ND DISTRICT, ARKANSAS

WASHINGTON OFFICE:
1330 LONGWORTH
HOUSE OFFICE BUILDING
WASHINGTON, D.C. 20515
(202) 225-2506

Congreſs of the United States
House of Representatives
Washington, D.C. 20515

DISTRICT OFFICE:
1527 FEDERAL BUILDING
700 WEST CAPITOL
LITTLE ROCK, ARKANSAS 72201
(501) 378-5941

October 31, 1979

The Honorable Carl D. Perkins
Chairman, Elementary, Secondary and
Vocational Education Subcommittee
House Education and Labor Committee
B346C Rayburn House Office Building
Washington, D.C. 20515

Dear Mr. Chairman:

Enclosed are copies of letters I received from Dr. James J. Hudson, Dean of the University of Arkansas Graduate School, and Carl S. Whillock, President of Arkansas State University, advising me of their concerns about the "truth in testing" bills, H.R. 4949 and H.R. 3564.

I understand both measures are before the Subcommittee for hearings, and I wonder if you would be kind enough to review and retain Dr. Hudson's and Mr. Whillock's correspondence as part of your permanent record of public comment on this issue. I'm sure their thoughts will prove valuable to the Subcommittee as hearings progress.

Your consideration of and attention to these views are sincerely appreciated.

Sincerely,

Ed Bethune
Member of Congress

EB:abb

Enclosures

Arkansas State University

OFFICE OF THE PRESIDENT
STATE UNIVERSITY, AR 72467
501/972-3030

October 1, 1979

The Honorable Ed Bethune
U. S. Congressman
House Office Building
Washington, DC 20510

Dear Ed:

Please give your most careful attention to the legislation proposed in H. R. 3564 and H. R. 4949 which relates to the standardized testing programs that are vital to American education. This legislation would seriously damage the systems of measurement and evaluation that are essential to our determination of the strengths and weaknesses of persons seeking admission to various higher education programs, the quality of our programs and their effectiveness, and the level of attainment in knowledge and ability by those people completing our programs and entering the many fields of employment throughout the American society. We strongly request your assistance in defeating these proposals.

Should you desire specific information about the proposals and the impact that would result from their adoption, we suggest that you contact the Council of Graduate Schools in the United States, the American Association of State Colleges and Universities, or the National Association of State Universities and Land-Grant Colleges, all of whom have offices at One Dupont Circle in Washington, D. C.

Your assistance with this important matter will be greatly appreciated.

Sincerely,

Carl

Carl S. Whillock
President

CSW:ph

OFFICE OF THE DEAN
316 Administration Building
(501) 575-4401
Fayetteville, Arkansas 72701

UNIVERSITY OF ARKANSAS · Graduate School

October 3, 1979

The Honorable Edwin Bethune
U.S. House of Representatives
Washington, D.C. 20515

Dear Congressman Bethune:

As Dean of the Graduate School at the University of Arkansas and President of the Conference of Southern Graduate Schools, I am concerned about the "Truth in Testing" legislation aimed at standardized tests. Although the intent of the proposed legislation appears unassailable on the surface, deeper analysis reveal reasons for concern.

One of the key elements of the Weiss bill (H.R. 4949) is the requirement for disclosure of all test forms--all questions on the test, together with the correct answers. This requirement is counter to the manner in which such tests as the Graduate Record Examination are developed in order to provide assurance that they are not compromised and to provide assurance that test results taken at different times and places have comparability. Great value is placed upon these features of the test and considerable effort in time, talent, and money are dedicated to this cause.

Standardized test scores represent one of the relatively few meaningful criteria for evaluating students for admission to colleges or graduate and professional schools. Admission decisions based on fewer criteria are more likely to result in unsuccessful experiences on the part of students, leading to frustration and disappointment and wasteful utilization of the resources by the institution. Consequently, every effective evaluation tool needs to be employed in the search for qualified students.

Although standardized tests have been criticized for possible cultural bias against minorities, the elimination of standardized tests would be detrimental to the minority student who has strong innate ability but who graduates from a less prestigious undergraduate school. Rather than taking action which will weaken or destroy standardized examinations, a more beneficial course of action would be to eliminate cultural bias from such tests. I understand the Educational Testing Service (the producers of the G.R.E. and several other popular standardized tests) is taking action to solve this bias problem.

The University of Arkansas is an equal opportunity/affirmative action institution

Page 2
Congressman Bethune

I strongly urge degeat of the Weiss bill and the similar Gibbons bill (H.R. 3564) and encourage support of necessary legislation to remove cultural and ethnic biases from standardized tests used in admission criteria for educational programs.

Sincerely,

James J. Hudson
Dean

ct

EMORY UNIVERSITY SCHOOL OF LAW
GAMBRELL HALL ATLANTA, GEORGIA 30322

OFFICE OF THE DEAN

TO: Congressman Carl Perkins, Chairman,
Subcommittee on Elementary, Secondary and Vocational Education.

FROM: Harold L. Marquis, Acting Dean

DATE: November 5, 1979

RE: Statement for entry into the record of the hearings on Federal Admissions Test Bills HR 3564 and HR 4949

The Bills currently under consideration pose a serious threat to the continuation of the practice of standardized testing of law school applicants. The administrative difficulties and legal pitfalls facing the test administrators under these Bills may well result in the abolition of the Law School Admissions Test. Not only would such a consequence result in the loss to most law schools in the United States of a proven valuable predictor of likely law school performance but also it would mean a return to the previous practice of evaluating large numbers of applicants for admission on the basis of undergraduate record and personal references alone.

The underlying assumption of these Bills seems to be that there are major defects in the pre-professional testing of applicants which need governmental attention. This assumption appears to be based upon a misunderstanding of the nature of the standardized test and the use to which it is put. The limited purpose of the LSAT is to provide Admissions personnel with a predictor of performance in a law school setting. As such a predictor it is highly useful and studies by ETS have demonstrated a fairly high correlation between LSAT scores and performance in law school for all applicants. Even so, it is not the only factor to be considered in the Admissions decision. Many factors go into that decision and no applicant should be admitted to law school unless his whole record including his LSAT score indicates that he is likely to be able to complete the law school program. Removing the only standardized predictor from this process would mean that Admissions personnel, faced with hundreds of applicants of varied backgrounds and experiences, would be more likely to make inaccurate comparative evaluations.

Thus the chances of admitting students unable to complete the law school program or denying admission to students who might well perform satisfactorily would be substantially increased. Moreover, the loss of the standardized predictor would be likely to work to the detriment of students having features in their pre-application background which make their credentials difficult to evaluate. For example, some students may have spent several years at home or out in the work force after completion of their undergraduate education. Others may have graduated from colleges whose reputation is not well known to the Admissions Committee. Without the assistance of the LSAT as a predictor of likely performance in law school, it will be extremely difficult to accurately assess the potential of those applicants as law students.

In sum, we at Emory Law School believe that the LSAT is a reliable and accurate device for predicting law school performance. It is not the only predictor but it has proven its usefulness. Had the Admissions Committee at Emory Law School been dissatisfied with the test or had reason to question its validity as a predictor, then its use would have been discontinued. We urge that these measures which may result in the loss of that predictor and the subsequent undermining of efforts to fairly and accurately evaluate candidates for admission to law school, be rejected by the Committee.

Harold L Marquis

Harold L. Marquis
Acting Dean

copy to: Congressman Wyche Fowler
Congressman Elliott Levitas

TESTIMONY IN FAVOR OF PASSAGE OF H.R. 4949

As a public school teacher and administrator for thirty years, I give my enthusiastic support to the passage of H.R. 4949, the Federal Truth in Testing legislation.

There is no question in my mind that students should have the clear cut right to examine the questions on which they are judged and see their corrected responses. Tests with this potential for control of lives should not be shrouded in secrecy—particularly when they are corrected by machines subject to mechanical error fed by attendants subject to human error.

I saw the possible devastating effect of standardized, machine scored test error first hand recently. When the medical school admission tests were administered to 27,300 students on April 15, 1978, 90 percent of them received scores lower than they should have. My son was one of those taking the test.

The error was found, however and medical schools were notified by the testing company to ignore the original scores—as were the students who took the test. While the net result was positive, "Three Mile Island" possibilities existed in the secretive process.

Test secrecy is perhaps the last major tyranny sanctioned by our Government. It should be lifted—and H.R. 4949 is a good vehicle for doing so.

Sincerely,

WILLIAM D. CORBETT.

APP 1HH-JPT

nln

national league for nursing · ten columbus circle · new york, new york 10019 · 212 · 582-1022

December 18, 1979

Hon. Carl D. Perkins
Chairman, Committee on Education and Labor
Subcommittee on Elementary, Secondary and
Vocational Education
Rayburn House Office Building, Room 346C
Washington, D.C. 20515

Dear Mr. Perkins:

The National League for Nursing is a coalition of nurses, other health professionals, consumers and health agencies -- 17,000 individual members and 1,800 agencies -- working together to help meet community nursing needs and to assure high quality nursing education. Among its many functions, NLN provides preadmission and achievement test services to state-approved schools preparing registered and practical nurses. The League also serves as the test service agency for the State Board Test Pool Examination for both registered and practical nurse licensure.

We would like to express our views on the Truth in Testing legislation that was the subject of hearings held by your Elementary, Secondary and Vocational Education Subcommittee last session.

We oppose the Gibbons and Weiss proposals for a number of reasons, not the least of which are the technical and administrative difficulties which would result from having to disclose test questions and answers to test takers. This provision, however, is likely to have the most hardfelt impact on testing candidates.

Methods of scoring and reporting of results would have to be changed; in addition to the extra time and staff resources that would be needed to compose new tests for each administration. The League has conservatively estimated that changes of this nature would cost approximately $100,000 annually.

a membership organization dedicated to meeting the health needs of the people by improving nursing education and nursing service

Yet, the brunt of the negative results that would occur from enactment of such a law would be on prospective test candidates. First of all, the previous flexibility of test agencies in administering make-up examinations and scheduling alternative dates for handicapped and religious Sabbath observing students will no longer be feasible because of the extra cost. This is precisely the situation which has occured in New York State with the passage of the LaValle Act.

The enactment of this legislation in New York State has posed other formidible problems. Presently, twenty of the twenty-six test programs threaten to be withdrawn from use in New York State. The main reason is that many of these tests are administered to small numbers of candidates and test agencies will not be able to absorb excessive new administrative and personnel costs. If these tests are withdrawn, the greatest damage will be to programs with highly competitive admission procedures. Loss of tests will not hurt the elite students from the most prestigious colleges; it is very easy for deans to select only the top students from the top undergraduate programs. Many schools review the applications of many other individuals who would not have been considered except for their good performance on the admissions tests. This group very often includes the economically disadvantaged student who was forced to attend a less prestigious college or the student who obtained lower grades in the first year or two of college because of a disadvantaged background. Although NLN has not discontinued our testing services in New York, as a result of the LaValle Act, we have had to raise our fees and decrease services to some extent.

We feel strongly that the institution of standardized testing should be upheld as a tool for the guidance and placement of applicants to educational institutions. NLN's preadmission tests have been used very successfully as predictors of future performance in schools of nursing.

Finally, we urge you to examine the severe financial repercussions and other disruptive consequences of the LaValle Act in New York State before considering this matter at the national level.

We would be happy to discuss this topic with you at some future date and answer any questions you might have.

Sincerely,

Pamela Maraldo
Public Affairs Director

PM:fw

New York State Department of Civil Service

THE STATE OFFICE BUILDING CAMPUS · ALBANY, NEW YORK 12239

COMMISSION
Victor S. Bahou
PRESIDENT
Josephine L. Gambino
James T. McFarland

John J. Mooney
ADMINISTRATIVE DIRECTOR

December 20, 1979

Hon. Carl Perkins
Chairman, House Subcommittee on
Elementary, Secondary and
Vocational Education
B-346C Rayburn Office Building
Washington, D.C. 20515

Dear Mr. Chairman:

In view of the conflicting and often confusing viewpoints being presented in the area of "Truth in Testing," it still seems important that those involved in setting policy or designing legislation should be aware of an information system in effect in New York State. The attached "Examination Review" is a summary of this system and highlights the procedures in use today. These have been developed over a number of years.

The system we use ensures that two objectives of testings are met: (a) The system is job related. Individual test takers can review the test materials (but they may not keep them) to determine whether the questions and answers are, in effect, job related. Following this step a second check takes place, namely, (b) the individual test takers can see and compare their answers to the correct or model answer, ensuring that the system has been applied correctly. This double checking would identify any mistakes made in scoring or transferring of scores, etc., made when the scoring system was applied to individual answer sheets. Individuals want to know whether they have been treated fairly and our system provides them with an opportunity to determine this for themselves.

You will note from the charts and graphs prepared as part of the statistical information we have gathered that the number of requests for this information is quite modest. Thus, it is possible that the issue, while an important one, has been blown clear out of proportion. If our experience in the employment setting generalizes to other settings, it would appear that only a few individuals will follow up and obtain the information desired, thus keeping costs at a minimum. In one of the instances, we cite only 3/4 of one percent of the test takers for a particular examination found fault and objected to any of the questions. Note that we do not give any test materials away, rather that these are reviewed in several locations under controlled conditions, and within a reasonable time period, after which the review opportunity is over and all materials are returned to us.

Let us know if we can be of additional help or if you would like any of the points clarified. I will be happy to talk with you, or you may contact our Assistant Administrative Director, Mr. John M. Keefe, for information on administrative concerns. He can be reached at (518) 457-6203. My telephone number is (518) 457-5358.

Sincerely,

Grace H. Wright

Grace H. Wright, Ph. D.
Research Psychologist
Director, Validation Program

Attachments

EXAMINATION REVIEW

The Department of Civil Service accepts the responsibility to maintain security and integrity in the examination process. This calls for fair treatment of all candidates and for controls to insure meeting these objectives. To make certain that no one candidate can benefit from unauthorized access to test materials, we enforce measures to protect those test materials at each step in the examination procedure.

After an examination is administered, all papers are returned from the testing centers to Albany for rating. When the papers arrive, they must be accounted for and checked for discrepancies. This is an important initial step in maintaining tight security and preserving the integrity of the entire examination process.

A rating key must be prepared for use in the preliminary scoring of answer sheets. The answer sheets are scanned for double answers, extraneous marks and messy erasures. If this is not done, scoring errors will occur.

In many examinations, we invite candidates to return at a later date to review the questions and the tentative key (correct) answers. (See Attachments.) At the pre-rating review the candidate is provided a test booklet and tentative answer key which are reviewed in the presence of monitors. The written objections received from all candidates who choose to make this review are analyzed in depth by the examining staff. In addition, the answer sheets are rated and candidates' responses to each question are tallied. Strict control is still exercised at this point to maintain exam security. Although legitimate candidates may review the questions and tentative keys, they must leave all copies of these questions and keys in the review center. This prevents leakage of the contents of the test to individuals who might take a similar test in the future in which the same or similar questions may be used.

The responses are reviewed, along with the candidates' objections, to determine whether there is reason to suspect the correctness of any tentative key answer. This review process also helps to reveal faulty or ambiguous questions and ensures that no candidate is harmed because a poorly constructed question was used.

The results of this review are transmitted with recommendations to a consultant. The consultant, in turn, analyzes the record, and may request information or discuss certain questions with the examining staff. The consultant then submits a report and recommendations to a committee of the Civil Service Commission. The recommendations of the committee are then considered and acted upon by the Civil Service Commission. This arduous procedure is designed to perfect the rating key which is then uniformly applied to all papers.

A frequency distribution is next prepared showing raw scores for all candidates. For example, the distribution might show that one candidate in a 90-question test had 84 correct responses, two had 81 correct responses, and so on for all candidates. This distribution form carries a great deal of data pertinent to the examination and to the positions to be filled.

The distribution form is analyzed and discussed by the responsible examiners. Where necessary, conferences are held with informed persons who can help reach decisions about the difficulty level of the test, the quality of the competition and the kinds of persons needed to perform effectively in the positions to be filled. The passing point in the examination is then set as is the formula to be used in converting the raw test scores to final scores. Not until final scores are computed do we initiate the procedure which reveals the identity of successful candidates.

The computational review takes place after the final answer key has been established. At this review the candidate is provided a copy of the final answer key and his own test answer sheet whose surface has been waxed to prevent tampering. The candidate may then determine whether the final answer key has been applied correctly to his own answer sheet and whether any other procedures, such as weighting, have been performed correctly. These materials also are reviewed under the close supervision of a monitor.

Post-rating reviews are a combination of the pre-rating and computational reviews. These allow candidates to review both the quality of the rating given their responses, as in a subjectively scored free-response type essay, as well as the accuracy of the calculations made in arriving at the final score. These procedures provide candidates with information about the test, about their performance on the test, and how their final scores were determined.

When the examining staff has finished its work, materials in most examinations must then be processed through computer equipment for preparation of the official list of successful candidates, notices to all candidates of the result of the examination, and other related documents. With the exception of analyses and other technical material which go into the permanent examination file for use in developing a new test, this completes the examination process as it relates to the candidates. The review results are filed with the other materials used in developing the test and are consulted when other exams are developed which may use the same or similar questions.

October, 1979
New York State
Department of Civil Service

JMK:GHW:CHT:sl

TABLE 1

Number of Examination Questions by Type with Rate of Appeals and Number of Examinations Serviced

Activity	1975-76	1976-77	1977-78	1978-79
Examinations held involving written tests (State, Local and OCA)	4,201	4,080	4,306	4,955
Test questions used	34,876	36,618	34,052	45,123
Questions from files - Not revised	25,894	28,807	28,259	35,904
Questions from files - revised	3,892	3,500	2,518	3,025
New questions - outside sources	1,975	1,790	1,611	2,975
New questions - staff written	3,115	2,521	1,664	3,237
Essay, design, drawing questions	301	75	33	44
Appealed multiple choice questions	2,440	2,139	2,469	3,147
% of multiple choice questions appealed	7%	5.8%	7.2%	7%

NEW YORK STATE DEPARTMENT OF CIVIL SERVICE
September, 1979

TABLE 2

Computational and Post Rating Review Activity by Month for a Six Month Period in 1979

1979

Jurisdiction	March		April		May		June		July		August	
	Comp.	Post	Comp.	Post	Comp.	Post	Comp.	Post	Comp.	Post	Comp.	Post
STATE												
Requests filed	734	5	637	1	423	2	435	1	560	0	579	0
Failed to appear	132	1	184	1	117	1	48	0	137	1	116	1
Object received	11	5	6	0	11	1	5	0	10	0	11	0
LOCAL												
Requests filed	227	5	123	9	100	9	225	4	123	6	135	10
Failed to appear	44	0	16	0	76	0	24	1	57	3	137	12
Object received	4	4	4	0	2	0	5	6	3	3	6	1

Note. These figures reflect activity in the Central Files Unit during given periods of time and do not necessarily inter-relate to each other on an exam series basis. For example, the 132 people who failed to appear for State computational review in March are probably not included in the 734 people who requested a review during March, nor can it be said that the 184 people who failed to appear in April are from that group of 734.

Computational review is limited to a check to see that papers have been scored correctly against a final key and does not allow appeals regarding the correctness of the key. Some appeals result from the candidate's failure to understand the scoring formula and not as a result of an incorrectly scored paper.

Post-rating review allows objections to the ratings given to a candidate's response (usually of the free response type-essay or drawing, etc.) as well as to the accuracy of the calculations made in arriving at the final score.

Examinations are held on about 15 scheduled dates a year (usually Saturdays) at 38 State and more than 100 local test centers. In recent years, more than 600 State and more than 3,000 local written tests were conducted annually. As many as 30,000 candidates have been tested in a single day. In 1978, tests for 220,000 candidates were processed.

NEW YORK STATE DEPARTMENT OF CIVIL SERVICE
September, 1979

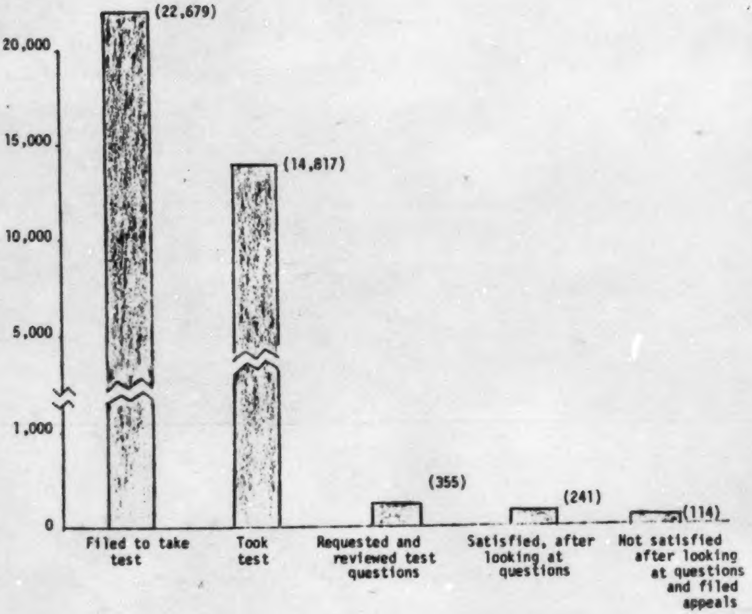
TABLE 3

Candidate Review Experience for a Recent Written Multiple Choice Test

Category	Frequency
Candidates filing for test	22,679
Candidates taking test	14,817
Candidates reviewing test questions and tentative key answers	355
Candidates not submitting objections to any tentative key answers	241
Candidates submitting objections to some tentative key answers	114

Note. A total of 330 objections were received for this 90 question test. Of these 90 questions, 61 received objections. Resolution is expected to take an additional 8 to 10 weeks. The large number of appeals was ascribed to the large and heterogenous nature of the candidate group and was not typical of exams in general. Even so, of those appealing (355), somewhat less than two-thirds (241) did not submit an objection after reviewing the test question.

NEW YORK STATE DEPARTMENT OF CIVIL SERVICE
September, 1979



FREQUENCY RATES

Figure 1. Candidate Review Experience

__Note__. Of the candidates who filed for a test (22,679), approximately one-third (7,862) did not show up to take the test. Of those who did (14,817), only 3/4 of 1% (.00769) have filed appeals, not being satisfied with the information provided during the review. This small group challenged sixty-one of the ninety questions in the test. (No demographic information is taken or kept of the characteristics of candidates who review items or of those who eventually file appeals.)

NEW YORK STATE DEPARTMENT OF CIVIL SERVICE
October, 1979